



**Facultad de Ciencias Económicas y Empresariales**

**TRABAJO FIN DE GRADO**

**GRADO EN ADE INTERNACIONAL**

**AN ANALYSIS OF SPORT SUCCESS IN EUROPEAN FOOTBALL.  
WHICH ARE THE MOST DETERMINANT PERFORMANCE  
INDICATORS TO BE SUCCESSFUL IN THE UEFA CHAMPIONS  
LEAGUE?**

Andrei Iulian Burlea

DIRECTOR

Fernando Lera López

Pamplona-Iruña

20 de Mayo de 2014

## QUOTE

“Soccer isn't the same as Bach or Buddhism. But it is often more deeply felt than religion, and just as much a part of the community's fabric, a repository of traditions.”

- Franklin Foer, *How Soccer Explains the World: An Unlikely theory of Globalization*, 2004  
(Harper Collins)

## ABSTRACT

The football or soccer is known to be the most popular sport worldwide becoming thus subject of different investigations. The aim of this study was centered on its performance analysis: identifying the most determinant game related performance indicators to reach success in European football, taking as research objective the UEFA Champions League teams during 2003 – 2008 seasons. Thus, any soccer stakeholder could increase its understanding about decisive playing actions to compete and improve their performance analysis on this competition. This short research was based on Opta Sports data including as criteria's 12 performance indicators and 4 goal variables. The results of the regression and ANOVA analysis showed that the most significant variables to achieve success were the scored goals, the realized shots, recovered balls, received goals at home and received shots on goal. Surprisingly, the possession is one of the least influential variables (together with committed fouls and red cards) despite recent focus on this way of playing promoted by teams as FC Barcelona or more recently Bayern München.

**Keywords:** *football, UEFA Champions League, performance indicators, statistical analysis*

## TABLE OF CONTENTS

1. INTRODUCTION .....	1
Objective.....	2
2. LITERATURE REVIEW .....	5
Performance Analysis .....	6
A recent study about UEFA Champions League .....	10
3. HYPOTHESES.....	11
4. METHODOLOGY AND DATA SET .....	12
4.1 Data Description .....	13
4.2 Variables .....	13
4.3 Multiple Linear Regression Model.....	17
5. STATISTICAL ANALYSIS .....	19
6. RESULTS .....	22
The descriptive analysis .....	22
The Correlation Analysis .....	27
The ANOVA Analysis.....	28
The Regression Analysis .....	29
The Confidence Intervals Analysis.....	31
7. CONCLUSIONS .....	34

## Table of Figures

### Variable Descriptions and Results

Table 1 : Selected variables for the study.....	15
Table 2 : Regression Model Independent Variables and their $\beta$ coefficients .....	18
Table 3 : Descriptive Analysis of UCL using average parameters observed on every phase .....	23
Table 4 : Correlation Matrix, Ranking vs. Independent Variables .....	27
Table 5 : Results from the ANOVA analysis.....	28
Table 6 : Regression Results from Figure 1 (Appendix I ) .....	30
Table 7 : Regression results including goal variables in the previous regression model .....	31
Table 8 : Average p-value of the 4 goal variables.....	34

### Graphs

Graph 1: Evolution of the two most objective attack variables .....	24
Graph 2: Evolution of the two most objective defensive variables.....	21
Graph 3: Evolution of the team`s goalkeeper parameters .....	22
Graph 4: Attack performance indicators.....	23
Graph 5: Defense technical performance indicators.....	23

## APPENDIX I

### Regression Model Assumptions: Testing Normality and Multicollinearity

Condition 1: Normality Test of Residual of the 12 Variables Regression Model.....	1
Condition 2: Normality Test of the Dependent Variable: UEFA Ranking.....	1
Condition 3: Linear Independence among Predictor Variables.....	1

### Regression Results from Gretl

Figure 1: Regression Results, Regression Model based on Table 2 .....	3
Figure 2: Linear Relationship, Ranking vs. Goal Variables .....	4
Figure 3: Linear Relationship, Ranking vs. Non Influential and Influential Parameters difference .....	5
Figure 4: Linear Relationship Ranking vs. Other Influential Parameters.....	6

## 1. INTRODUCTION

Football has become the most popular sport worldwide in the last decades taking into account criteria's such as TV audience, participation or revenue earnings figures from all stakeholders (Palacios-Huerta, 2004).

It is commonly known that most of football teams are adored in their home cities and this makes football different from most of sports in terms of appeal and fans support.

This phenomenon is even more surprising when observing that star players are more famous than religious and political leaders; when sport newspapers having mainly football contents are the best-selling in European and Latin American countries.

To observe the huge effects and trend of this sports let's have a look at the audience figures: during 2010 World cup in South Africa the average official rating 188 million people per match and there were 700 million people watching just the final, beating the previous audience record broken by the 2008 Beijing Olympic Games Inauguration ceremony (Associated Press, 2011). This is a clear evidence of audience success and it has huge impact on population worldwide.

A part of being followed by billions of people, it has a strong participation as well being represented officially in 208 countries (FIFA, 2014). There is no similar organization being so strong globally and having such impact on its followers or consumers.

Looking from an economic-marketing point of view, many corporations found football as a unique and perfect opportunity to increase their brand awareness through sponsorships or advertising during football events.

Due to all these previous figures we just saw, football was even classified as a social phenomenon capable to influence mass of people and it was subject of several investigations. It couldn't be otherwise as there was a strong interest to understand what makes this sport so attractive (Dobson & Goddard, 2011).

Social sciences investigated football since 1950's (Dobson & Goddard, 2011) from various perspectives: as a game, as an industry and as a product with a huge demand worldwide. The

revenues structure of the industry investigation was made by Atkinson, Stanley & Tshigart in 1988 followed by the demand research done by Peel & Thomas (1992) or Kuypers (1995).

However, even if it was subject of many investigations; there was a limited research done in one of the most important aspects of this sport: the game outcomes. Which are the most important factors to reach football success and consequently achieve more attractiveness?

## **Objective**

The aim of this paper is then to identify what makes teams to be successful when competing on the pitch, choosing to analyze the UEFA Champions League. Which are the game related indicators, which are the first determinant steps the teams follow to reach success when playing football in this prestigious competition? This paper will answer which are the most determinant football actions that made teams to reach a top position in the biggest European club competition for the 2003 – 2008 seasons.

Furthermore, the results found out in the study will be compared to previous findings speaking about football performance. This will give the reader a general idea of different points of view about the performance analysis in football and comparisons will be applicable.

Most of teams are adored either because of winning or because of having a certain way of playing that makes their fans feeling proud of it. There are even football philosophies that have opposite patterns of playing and this needs more research in order to understand this sport's attractiveness. An example of this is one of the biggest rivalries in this sport nowadays, in 2014, the one between FC Barcelona and Real Madrid CF; their fans adore these teams and they have a special way of playing: possession and passing vs. ball recovery and goal efficiency. Which is more efficient? We can say before analyzing anything that both of them are, but which is more determinant and representative in UEFA Champions League? Those are the questions will be answered later on the results section.

One can ask himself about why studying the success determinants related to the way of playing while this sport has already a clear economic and social success. The study is focused on those performance indicators basically because the most followed teams globally are the ones having a singular way of playing and at the same time they reached remarkable success in past football events due to winning on the pitch with that style. I mean that the way of playing could be

many times one of the first triggers of increasing success; TV audience and fan support from thousands to billions of people during a century of football history.

After reaching a global importance and being a social phenomenon (Palacios-Huerta, 2004), it is clear that football deserves a careful study of its most determinant game-related factors. As everything starts from winning by playing on a pitch, success starts from that too.

Consequently, as there's a scarce investigation on technical performance analysis we're going to focus on that instead of the goal variables. In any case, the goal variables influence will be analyzed as well, but they won't represent the main objective of this paper.

## **Methodology**

To separate and find out the specific influential variables on the team's success there were performed a simple and a multivariate statistical analysis. The data review starts with a descriptive analysis is going to compare the teams from the 5 UCL stages: group stage, eight-finals, quarter-finals, semifinals and the finals. Their average scored and received goals, the possession and other related variables will be compared to see the most outstanding differences.

Then, a correlation analysis is performed to see the relationship between variables, and specially the relationship strength of predictor variables to the success variable: the overall ranking. Moreover the Correlation will help us to eliminate the independent variables that are highly correlated in order to avoid multicollinearity while building the Regression Model.

Afterwards an ANOVA analysis will be performed especially to find out the p-values of every variable, including just here the goal parameters. The results of this study will be compared with the previous and the coming ones. Next there will be performed a Multiple Variable Regression analysis: a regression model will be build based on the chosen performance indicators and on the basic assumptions as well: homoscedasticity and normal distribution of error variable and linear independence between predictor variables.

At the very end of our data review there is going to be performed a confidence interval hypothesis testing analysis for the regression statistics we obtained using the F or T-statistics and the p-value as criteria's in order to accept or reject the hypotheses will be posed later on.



All these analyses will be performed using the Gretl statistics software. After introducing manually every team's value for the 16 chosen variables in Excel there has been created an adequate database that was updated to Gretl. Once it is imported any user can perform descriptive or any other statistics related analysis.

As it was said before as well, the analyzed competition for the study was the UEFA Champions League, one of the most prestigious football clubs tournaments nowadays; known for joining the best European clubs every year. An example of its prestige is the last year final which has been followed by 360 million people worldwide (Ashby, 2013). Many times, the most successful European clubs are compared using the number of UCL<sup>1</sup> trophy's they have won which makes it one of the most appealing tournaments to analyze as it is a synonym of success.

All primary quantitative data used for the statistical analyses corresponds to Optasports Database ([www.optasports.com](http://www.optasports.com)), a known sports data company providing data useful for performance analysis in sports such as rugby or football. It has representative offices in many European and South American capitals. Some of its clients in bet, communication or sports industry would be: Bet Fair, William Hill, Sky Sports, ESPN, The All Blacks, Chelsea F.C, Manchester City and Real Madrid. We can see then it is a quite trustful database that can be used for multi-sport performance analysis. I had the opportunity to use all necessary data after a collaboration agreement between UPNA and Optasports.

The data pack incorporated the five years seasons between 2003 and 2008. What Optasports provided was annual data for every team, so that I disposed of 12 variables in annual terms: total shots on a season, total centers made on a season, total yellow cards in a season, etc. Moreover, there have been created 3 variables based on official data available on the UEFA website: Overall ranking in the competition; number of played games and the average possession during the whole competition.

The results of this study will be helpful to increase the understanding of football stakeholders about the most distinctive performance indicators when participating in UCL. So that, football debates can have a slightly empirical approach when affirming for example if scored goals objective is more important than the defending goals one on a team's overall success; or that

---

<sup>1</sup> UCL is the abbreviation of UEFA Champions League

the possession should be the main factor to measure a team's performance. Also, we will be able to give an answer to the question if possession is more effective than counter-attack (ball recovery) style; all these questions applied to the 2003-2008 seasons taking into account just the data of the 160 teams participating during these seasons. Many similar questions will be answered below.

## **2. LITERATURE REVIEW**

The football's success was well known more than 100 years ago but the research of this phenomenon started after the 1950's (Dobson & Goddard, 2011). The most commonly analyzed football events were usually local leagues like Premier in England; La Liga in Spain or the World Cup (Collet, 2013; Lago – Peñas & Lago – Ballesteros, 2010; Atkinson et. al, 1988).

Trying to investigate a different top competition, the intention of this study is to analyze the most determinant factors that make teams win and achieve success in UEFA Champions League. Which are the most important factors to reach a top position?

As it was said before, the realization of this study was possible due to the Optasports and UPNA agreement. Those were the primary sources used for the statistical analysis together with many secondary sources as articles from sports journals or statistics books were consulted as well. It was interesting to see the football performance findings of many authors and their different interpretations. Later on we'll be able to see there is diversity within their interpretations mostly because the analyzed competitions and the utilized parameters were different; so it could be said that it was quite improbable to come to similar conclusions starting with different approaches.

Before continuing with the study, let's make an introduction to performance analysis and the most used parameters in previous studies.

## **Performance Analysis**

Even if soccer counts with an important fan base and worldwide audience; there have been limited studies on its game outcomes. There are many studies about its economic and social impacts (Dobson & Goddard, 2011) but few of them related to its way of playing. Starting with 1950`s, many authors started to be concerned about the impact of this sport and specially about the performance analysis of the football. Which is the normal pattern of football performance analysis?

The performance analysis is usually performed using statistics taking into account past records of performance indicators. But, what is a performance indicator? A performance indicator can be defined as a “selection, or combination, of action variables that aims to define partially or completely a performance in a certain sport “. In order to be efficient; any performance indicator should be evidently related to success. (Hughes & Bartlett, 2002) (Rumpf, 2014).

The most previously investigated action variables could be categorized into physiological, technical and tactical ones.

- Physiological Parameters: are the variables related to physical capacity of players (Carling, Bloomfield, Nelsen & Reilly, 2008) (Lago-Peñas, Rey, Lago-Ballesteros, Casais & Dominguez, 2009)
- Technical Parameters: factors as shots on goal, realizing centers, dribbling's , ball recovery and tackles or passes (Rumpf, 2014)
- Tactical Parameters: ball possession; offensive , transition and defensive style of playing (Rumpf, 2014)

If we would make a list of the most relevant performance indicators from previous studies, that ranking would include:

- Scored Goals: (Lago-Ballesteros & Lago-Peñas, 2010)
- Total Shots on Goal: (Lago-Ballesteros & Lago-Peñas, 2010)
- Goal Efficiency (Goals/Shots on Goal) (Lago-Ballesteros & Lago-Peñas, 2010)
- Possession (Collet, 2013)
- Passes Accuracy (Redwood-Brown, Bussell & Singh, 2012)

From all these parameters, even if physiological parameters have a clear influence on the game outcomes (Carling, et al., 2008), this study is going to focus more on the technical and tactical factors as shots, passes, area centers, scored and received goals. Because of available data limitations there are variables such as successful passes that will not be taken into account even if it's clear they have a strong influence on any team's performance.

As we're here to comment on previous football research, it is common to see many articles where the performance studies about football outcomes focus on analyzing the physiological estimates or individual player performance rather than technical team performance indicators, being there a clear research scarcity especially on team success analysis (Lago-Ballesteros & Lago-Peñas, 2010).

There are other additional observations telling us that many of previous and actual studies consider mainly relative tactical and individual performance estimates studies (Svensson & Drust, 2005; Di Salvo et al., 2007). The soccer players are analyzed individually according to their position and physiological form. We can understand then that the approach of these studies is more to find out better ways of training and improving individual players than a team's success. It is not our objective though.

Returning to a previous idea of tactical performance indicators, authors like Hughes & Churchill (2005) tried to observe playing patterns of successful and unsuccessful teams resulting in shots during the 2001 Copa America competition. They finally found out that there were no significant differences between the observed teams playing patterns leading to realized shots. (Lago-Ballesteros & Lago-Peñas, 2010)

Observing patterns of playing such as central zone or wing attack; having a defensive or attack style and attitude do not help getting clear ideas of measurable parameters of a team's performance as a whole in its participations.

Considering other aspects, it can be hard to extract a general conclusion about football performance because many authors focused their study in just one parameter, being it either possession (Collet, 2013) or the physiological form of players (Lago-Peñas et al., 2009). We know that football is composed of many parameters (not just one) and they should all be taken

into account when trying to measure success; that's why independent performance indicators would be a better and a more an empirical way to measure of a team's performance.

There can be another decisive factor that should be taken into account when speaking about all previous studies. Most of them analyzed a limited number of teams. Having a small sample by analyzing just a local league during a season would count just for a total of 20 teams measuring just a local cup can analyze just around 20 teams as well; that's why those findings cannot be applied to all football in general and that leads to conclusions which can be seen as relative.

After observing these playing patterns, individual and physiological approaches, it can be said that there was a need of more quantifiable and broader measures that can be taught to all professional coaches so that they can train the most important aspects of team game as football. That's why, some authors started to focus more on teams technical performance indicators as goal efficiency, shots, passes, possession, etc. (Hughes & Bartlett, 2002). The main reason seem to be that they can offer a more objective way to measure the performance of competing teams and there can be extracted interesting forecasts about future success of certain teams depending on their observed performance indicators.

Even before competing, coaches can focus on training the variables that are more determinant in a given competition instead of spending time on training the ones that do not have a strong impact on the team's success. Even so, there are many other variables that can influence a team's performance such as players experience, type of competition (league vs. knock-out system) or match location (home vs. away) but we'll focus just on a team's technical factors.

Furthermore the technical performance indicators can be useful to make comparisons between teams from different leagues or other competitions: which team has higher scoring goals efficiency? This type of questions can give a coach a clearer answer about the tactics to follow before playing against its opponent rather than just knowing its rival team is defensive.

There are also studies making comparisons speaking about team's performance indicators as possession, goal efficiency or defense importance (Collet, 2013). Sometimes their interpretations are contradictory and in this paper there is going to be added few more related conclusions to all previously analyzed performance indicators in UCL.

Example of team technical performance indicators: goals, total shots, total centers, recovered and lost balls, yellow cards, etc.

Taking the variable possession as an example; most of authors speaking about Spanish and English local leagues (Lago-Peñas & Dellal, 2010; Collet, 2013) concluded possession is relevant among successful teams as the observed successful teams (the champions or top 4 teams) used to dominate it. Nevertheless, there are other authors like Bate (1988) and Stanhope (2001) are not in line with those findings and they conclude that possession is not related to success of national teams.

We were before speaking about limited samples, but there's an example of a wider study that analyzed and compared all previous possession researches (Collet, 2013), covering many national and international competitions. This study is different to the previous ones as it has more representativeness (it counts with different profile teams and competitions) and its conclusions can be applied to football in general, but it has the limitation of using just a parameter when analyzing a team's performance: the ball's possession.

Collet (2013) found that possession is more relevant in national leagues than when competing in international club competitions like UCL (league and knock-out system). In the national teams competitions organized by FIFA the importance and significance of possession over a team's success was unclear yet.

Here again, we see different interpretations and this suggests us that there is a need for even more research and specially a broader one. Most of previous studies covered as maximum 3 years (Collet, 2013) and that's hard to get a general and homogenous understanding of football success determinants. It can be said that it is normal to get different interpretations as the previously named authors analyzed different competitions having different profile teams and structure: national vs. private clubs; league vs. knock-out system.

In any case there is a need for a joining measure, that unifies all success determinants criteria's and the performance indicators we introduced at 2.1 would be a solution. Counting with such indicators will basically help us to make comparisons between teams and a measurable contribution to success of every team will be able to be performed. Later on it will be discussed the results of this study made on possession influence over success in UCL.

## **A recent study about UEFA Champions League**

As it has been said before, UCL is the most prestigious football tournament for clubs in Europe and one of the most prestigious worldwide being watched by hundreds of millions people but there haven't been made too many studies about team performance indicators about the competition. Such a tournament shouldn't be missed when analyzing the top teams from all European countries. That's why this study is focused more on Uefa Champions League considering the scarce research that has been done about it.

Nevertheless, there is a recent study done by (Lago-Peñas et al., 2011) analyzing the differences in performance indicators between winning, drawing and losing teams in UCL for the 2007 – 2010 seasons. At a first sight, it seems a broad study that covered many seasons and it used different performance indicators; making it e a complete football performance studies applied to UCL.

Their conclusions state that the most distinctive team performance indicators were total shots, shots on goal, passes, successful passes and possession for winning teams, whereas for losing teams the most distinctive variables were yellow and red cards. Many of these variables are included in this paper as well and this is an interesting opportunity to make a comparison to see if some of the final results serve to reach similar conclusions.

Before starting to explain the methodology, as this short research has a similar approach considering the competition and the performance parameters with the Lago and Ballesteros (2011) study, there can be done a structural comparison. The covered period of the present paper is broader and prior (2003-2008) to Lago and Ballesteros's study (2007-2010). Just because of these similarities it's going to be interesting to see if the final results will coincide when considering similar measuring variables of success. We will be able to affirm and compare if the same variables as total shots, possession or yellow cards are consistent in a team's success over the passing of years.

That's why I'm thankful to the research done by all previously named authors because by reading their diverse studies I could identify the way the research about football performance analysis has to follow in order to improve.

### 3. HYPOTHESES

Our main conclusions will be based on the results of an ANOVA and regression analysis. How is the regression model going to be used on the statistical analysis? The main intention of creating a regression model was to make some hypotheses testing and confidence intervals for the regression coefficients. That's how, we'll be able to reject or accept some of the hypotheses we may pose.

One can also think himself that, there should be made a clear estimation of the effect each parameter has on the ranking. This is not the main objective of the present study, being just the identification of the most relevant variables over the ranking the aim of this review. Based on the literature review there will be posed three different hypotheses that will help us to get some ideas from the analyzed data.

- The main hypothesis, H0: The 12 performance indicators do not affect the ranking

After observing every variable's p-value this paper will be able to state if that is true or not. This is the hypothesis related to the main objective of the study, through accepting or rejecting it we'll see the performance indicators that most influence a team's success using criteria's as the p-value and confidence intervals for the t-statistic of every variable.

- A second hypothesis, H0: The possession is not determinant on a team's success

The results of this hypothesis will be compared to the ones found by C. Collet (2013), which affirmed that the effect of the possession on a UCL team's performance is low compared with other competitions as national leagues.

- A third hypothesis H0: received goals are more determinant over performance than the scored goals

Considering the third hypothesis, this is the only case we are going to consider the goal variables as it was the main reason to include these variables into the study. There will be performed an ANOVA analysis to observe all variable's p-values and then get the relevant conclusions.



At a first sight, we may clearly say that of course scored goals are more important. But this question is actually trying to respond to another question: which of both variables is more significant on a team's performance in UCL? Is it better to have an attack or defense strategy to improve your team performance? That's how, we'll observe their p-values and we'll see which one has stronger influence on a team's performance. Moreover this will permit us conclude if attacking or defensive teams use to be more successful.

The three hypotheses will be contrasted observing the regression ANOVA results and a Gretl Table results with every variable's relevant coefficients: Beta, t-statistic and p-value.

## **4. METHODOLOGY AND DATA SET**

The methodology used to carry out this study corresponds to a descriptive, dependence and variance analysis of the collected data.

### Steps to finish the study

- 1<sup>st</sup> Creation of a new Database inspired on Optasports that comprised all teams
- 2<sup>nd</sup> Select the relevant variables based on the literature review
- 3<sup>rd</sup> Start a statistical analysis taking into account the created database and previous studies
- 4<sup>th</sup> Start the development of this study

As it was introduced before, I was able to use the necessary dataset for this study after due to an agreement between UPNA and Optasports ([www.optasports.com](http://www.optasports.com)). In each team's registers there were a wide number of variables recorded such as scored goals, scored away and home, scored goals off the set pieces, scored goals from individual moves or scored goals from penalties.

For every variable there were observed similar details, even if there is an important variable missing from the database: the passes. Many researchers use either the passes or possession to analyze the same thing: how long does a team has the ball when playing; so it can be understood that the possession is going to measure the passes accuracy.

After observing all variables, there was made a selection of the most common ones and they were reintroduced in a new Excel database organized by seasons going from 2003 to 2008. This way we could have a unique database for all teams and all variables adding new parameters as the reached phase during the tournament; the country of origin of a team or the number of played games. The main reason was that there was a need of organizing all the information in a numerical format so that the Gretl software would permit performing all the necessary statistics analysis. That reorganization was the first step of this research, and then all the following steps could be correctly done.

#### **4.1 Data Description**

Normally it should be checked if there was taken a representative sample; in this case there were analyzed all participating teams during a period of five years, which counts for  $n=160$  observations. There were 32 teams per season and the competition observed phases were: the group stage; the eight-finals; the quarter-finals, the semi-finals and the five finals. The competition has a double structure: it is played on a league format during the group stages and on a knock-out system from the eight-finals on; just the top 2 teams of every group are the only ones that can advance to the knock-out system. The five years data was a result of the observations of around 1250 played games and every team played on average 7-8 matches.

All these figures, inspire to tell this paper's conclusions will be based on the total UCL "population" covering the 2003-2008 seasons and these findings shouldn't be biased due to the sample size. Compared to previous results, it can be said that this is a wider study if we take into account the sample size and hopefully it will serve to contrast the final results with Lago Peñas's (2011) and other previous similar studies.

#### **4.2 Variables**

Once we had as objective to find the success factors, one can ask himself about which would be the main variable, the one measuring success? I thought about three alternatives even if there could be many others. There are even professional coaches that choose possession or received goals as success variables, instead of won trophies.

##### **Scored Goals and Received Goals**

This could be a relative success variable that would show the strength of a team scoring or avoiding receiving a high number of goals. It would be relative because we know even before making any analysis that goals are very determinant when measuring success, so we couldn't get any relevant conclusion about how teams should play to improve their performance while competing in UCL. The answer would be, "yes", goals are synonym of success but how to

improve and start scoring more goals? In that case we need to think about other variables as well.

#### Number of played games

This is an own variable, created after observing the reached phase of every participating team from the UEFA website (es.uefa.com). The ones that couldn't pass the group phase played just 6 games, the ones that reached the eight-finals and quarter-finals played 8 and 10 games respectively; being the semifinals and the finals the synonym of success after having played 12 or 13 matches.

So the symbol of total success would be playing 13 matches in a season. This could be a more interesting option than the previous one and we could explain which variables influenced teams reaching a top final phase.

#### Overall Ranking in UCL

Compared to the national leagues, in UCL there's no such team classification with the champion being at the first position and the less performing team being the last one in the low tier. That's why initially it was difficult to measure the performance according to ranking.

The solution came from the UEFA webpage, which has a statistics record for individual players but also teams since 1955. From there (UEFA, 2014), it was extracted the annual overall classification of the 32 participants according to criteria's as scored and received goals, possession, total realized and received shots, etc. The 2 finalists were in the first position; the 2 semifinalists were in the 3<sup>rd</sup> and 4<sup>th</sup> position and so on. The teams having the lowest estimators on scored goals, realized shots and so on, were the ones being in the last positions; so the order made a lot of sense.

After finding this classification, it was easier to choose the dependent variable for our regression model, which it was going to measure how technical parameters as shots would affect the final position in the overall ranking. A previous interpretation would make us think that being in top position would imply having a high values in attack parameters and lower ones in the defensive; later on we're going to contrast it as well.

In the end, the final choice to measure team's success was Overall UEFA classification as it takes into account most technical parameters compared to the other two variables.

### **Variables transformation: from seasonal to per match parameters**

As it was explained in the sample description, all variables were explained in annual terms without taking into consideration the number of played games.

In order to make an accurate comparison between teams it was needed to transform those annual/seasonal variables in per match variable, so that teams didn't reach the final of tournament could be compared to the ones that did in terms of match variables: goals/played games, shots/played games. This way we could see the real efficiency of every team counting the number of games they played. By making this transformation it can be got a more realistic view of team's performance analysis. Before it was obvious that teams which reached the final had higher values in some variables due to playing more games so it is an example is any team having higher number of realized shots on total because of playing longer. Transforming the seasonal variables into match ones permitted to eliminate that drawback.

### **The dependent variable**

The dependent variable, the one explaining a team's success in UCL is the Overall Ranking taken from the UEFA historical statistics index (UEFA, 2014). The next table is going to take into account all variables: the dependent one, the predictor variables available from the Optasports Database and the other ones created using the UEFA website. From the total variables presented on **Table 1** there will be discarded the four goals variables and the number of played games when doing the regression analysis. Nevertheless the same variables will be taken into account for the other statistical analyses.

**Table 1 : Selected variables for the study**

Variable	Source
Overall UEFA Champions League Ranking	The dependent variable for the Regression Analysis , created from UEFA website
Number of Played Games	Created Variable from UEFA website
Scored Goals	Optasports Database
Scored Goals Away	Optasports Database
Received Goals	Optasports Database
Received Goals at Home	Optasports Database
% of Ball Possession	% of ball possession from UEFA website

Variable	Source
Realized Shots Per Match	Optasports Database
Realized Centers Per Match	Optasports Database
Committed Fouls	Optasports Database
Lost Balls	Optasports Database
Recovered Balls	Optasports Database
Received Shots	Optasports Database
Received Centers	Optasports Database
Received Shots on Goal	Optasports Database
Goalkeeper Saves	Optasports Database
Yellow Cards	Optasports Database
Red Cards	Optasports Database

Source: Own research; Optasports Database

### Independent Variables

All Optasports variables are technical performance indicators; they are a perfect measure to make comparisons to similar studies and their findings as many of them were previously analyzed: the possession is an example of them. The only two new variables included, were the scored goals away and the received goals at home. I thought it would be interesting to include those two variables as in UCL those scored and received goals count double when scoring or receiving; so they were a clear measure of distinctiveness among teams. If a team is scoring away and is not receiving many goals at home and keeps competing theoretically it could be more successful than if it's scoring just at home and moreover it receives many goals at home. Moreover, there were two variables taken from UEFA website: the ranking and the % of ball possession (the main reason was that it was a more consistent variable<sup>2</sup>). A third one was created so that to unify the way of measuring all variables and not to discriminate the teams that reached just the first phase of the tournament: the number of played games. That's how a team's performance could be measured depending on the played matches and not on the total parameters achieved compared to the champions.

Just 13 of **Table 1** variables will be selected to build the regression model: the overall ranking and all technical parameters excepting the goals and number of played games variables (hypothesis 1). The main reason of making this choice is to be coherent with the objective of this study that was identifying the technical parameters; that's why there were selected the

---

<sup>2</sup> The possession under Optasports criteria was explained in minutes; when it was converted in percentage it didn't give a realistic result compared to all other sources as all possession was similar and very close to 30 % for all teams estimates. That's why UEFA possession average parameter was considered as more realistic/actual and it was preferred to the Optasports one.



variables fulfilling those criteria's and they are shown later on **Table 2**, after the regression model is explained.

Even so, there will be made an ANOVA analysis for the goals variables together as we have it as a third objective. We'll try to identify their p-values and see which goals factor has a stronger impact on a team's performance (hypothesis 3). That can be useful to see which of the attack or defense focus leads to success.

### 4.3 Multiple Linear Regression Model

The Regression Model will be in charge of explaining the effect of the technical playing parameters on the final ranking. In order to build the regression model, there was consulted a statistics book (Newbold, Carlson & Thorne, 2010) which permitted to identify some relevant assumptions in order to create a reliable model. Next these assumptions are explained and there will be a representation of the final regression model.

**When is the Linear Regression Model estimating correctly the effects of the predictor variables on the dependent variable?**

- There is a representative sample  ; this study used the 100 % population of teams in that period, so this assumption holds.
- Predictors are linearly independent  : not sure 100 %. This condition is very important to measure the marginal change in the independent variable; a good measure of marginal change in ranking is only possible if the predictor variables are not strongly correlated. We will not do that kind of estimations, but it is better if this assumption holds. After observing the correlation matrix between the chosen indicators we will have to eliminate the variables with a high correlation coefficient from the model ( $r$  more than 0, 85) in order to avoid Multicollinearity<sup>3</sup> (Newbold, Carlson & Thorne, 2010). That happens especially with the variable number of played games and the goals variables: scored goals at home are clearly related to total scored goals. For that reason, we'll remove the highly correlated variables from our Regression Model. And we will take into account just the

---

<sup>3</sup> Multicollinearity is a statistical phenomenon in regression analysis where various predicted variables are highly correlated and one of the predictor variables can be explained by the other". In our study, the total scored goals is partially explained by the scored goals away for example.

technical parameters related to playing on the pitch. The goals don't give us clear ideas of how to improve performance, anyone knows that scoring goals helps to reach to success; but what it is more important to determine, is to see what actually and firstly drives to score more goals and then reach success.

- The variance of the error is constant (homoscedasticity) ✓ : Gretl gives us the option to create a model correcting the heteroskedasticity<sup>4</sup> so that this problem can be corrected anytime.

If these assumptions are followed when building the regression model, the estimations will be appropriated and they will better explain the relationship between the predictor and the dependent variable. Moreover the hypothesis testing and confidence intervals will be more reliable. So that's how we come to the final multi variable regression model, taking into account just the technical parameters. Next the coefficients and symbols of every variable are being explained.

**Y<sub>i</sub>**= Overall Uefa Champions League Ranking; this is the dependent variable and its values go from the 1<sup>st</sup> position to the 32<sup>nd</sup> one, representing 32<sup>nd</sup> the worst position in a season.

**i**=number of observations  
**B<sub>0</sub>**= Constant, this variable would be the position in the ranking without competing but we consider it has no relevance in this study and we'll study the

competing teams just. **ε<sub>i</sub>** = error term .In the next table there will be introduced the symbol of every variable and the final regression model explained as a linear function.

**Table 2 : Regression Model Independent Variables and their  $\beta$  coefficients**

X1, measured by $\beta_1$ = % Uefa Possession	X2, measured by $\beta_2$ = Realized Shots	X3, measured by $\beta_3$ = Realized Centers	X4, measured by $\beta_4$ = Committed Fouls
X5, measured by $\beta_5$ =Lost Balls	X6, measured by $\beta_6$ =Recovered Balls	X7, measured by $\beta_7$ = Received Shots	X8, measured by $\beta_8$ =Received Centers
X9, measured by $\beta_9$ = Received Shots on Goal	X10, measured by $\beta_{10}$ = Goalkeeper Saves	X11, measured by $\beta_{11}$ = Yellow Cards	X12, measured by $\beta_{12}$ = Red Cards

<sup>4</sup> Heteroskedasticity means that the standard deviation of a variable is not constant. In order to have a reliable regression model and get unbiased interpretations about it is needed to have a constant standard deviation within the independent and the error variables (Homoskedasticity). Gretl gives that option when treating a set of variables.

*Multiple Linear Regression Model:*

$$Y_i = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_4 * X_4 + \beta_5 * X_5 + \beta_6 * X_6 + \beta_7 * X_7 + \beta_8 * X_8 + \beta_9 * X_9 + \beta_{10} * X_{10} + \beta_{11} * X_{11} + \beta_{12} * X_{12} + \epsilon_i$$

In order to perform the regression analysis there will have to be made a normality and a collinearity test between all predictor variables each assumption holds. These tests are reflected on **Conditions 1-3** from the Appendix and they show that the model satisfy all necessary assumptions in order to get unbiased estimations and interpretations about the variables that explain the success in UCL.

## 5. STATISTICAL ANALYSIS

After introducing the methodology and the data used; this is the time to explain every step it was made to identify the most influential performance factors over a team during the 2003 – 2008 observed seasons. There were made five different statistical analysis and next we're going to give more details about how were they realized.

### Descriptive Analysis

The first step was to make a descriptive analysis among every phase and the team's characteristics: average scored and received goals per match, goal efficiency and defensive weakness among teams representing the different stages (groups, eight-finals, quarter finals, semifinals and finals). By observing that data, we could create some useful graphs to see the evolution of a selected group of parameters and that would give us an idea of which factors could have been determinant to reach another phase. For this description there were created two more variables: goal efficiency described as scored goals divided by the realized shots; and the defensive weakness defined as received goals divided by received shots. This analysis is made directly with Excel by grouping the data from every competition phase and then start inserting the most interesting variables graphs to see their evolution over the course of the tournament.



## **Correlation Analysis**

This Analysis was made to get a first impression of the relationship intensity and direction between the independent and the dependent variables. The analyzed variables were all the per match transformed variables except the number of played games. Using Gretl, there was executed a correlation matrix to obtain all the coefficients.

## **Regression and ANOVA Analysis**

The Regression analysis is performed using the multi variable model described in the **Table 2** (previous page); there were selected the twelve independent technical variables explaining the success dependent variable overall ranking in UCL. As most of assumption hold, the results of the Gretl analysis shouldn't lead us to any biased conclusion. There would be a last problem to address to the model, the homoscedasticity assumption (Newbold, Carlson & Thorne, 2010); but Gretl software has a tool to correct it as well and this is one more reason to trust on the results will be found out of this analysis. So there was taken all data from Excel and imported to Gretl, next there was executed a Homoskedasticity correction Linear Model and we got the necessary results for our regression Interpretations. There were other alternatives as OLS but I considered this was the best option having the most realistic and unbiased results.

The one way ANOVA analysis is made differently; here there are additionally included the goals variables as well as the 12 independent parameters in order to find out their p-values. We want to see first which of the goals variables have the lowest p-value; representing that the strongest influence over the ranking; and secondly to see the p-values of the same performance indicators as the taken ones in the regression analysis. The ANOVA is performed through Gretl as well; there is taken one explanatory variable against the ranking in order to observe the predictor's variable p-value. That value is going to be compared with the found one in the regression analysis to see if there is coherence in the conclusions. That way the study would have double and a more complete vision about the lowest p-values would make us reject the null hypotheses.

## Confidence Intervals and Hypothesis Testing

Once it has been seen the effect of variable's influence over team's success, we can now create the necessary hypothesis and contrast them using the F or T-statistics and the p-values criteria's.

- The first hypothesis H0: The 12 performance indicators do not affect the ranking
  - Is it true that the selected variables affect the ranking position in UCL?

H0: Predictor variables do not influence the final position in the ranking

$$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12} = 0$$

H1: Predictor variables can influence the final position in the ranking

$$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12} \neq 0$$

Significance level: 0, 05; Confidence Level: 95 %

Confidence Interval for the H0:  $(-1, 96 < t \text{ or } F < 1, 96)$

In the case that observed F or t-statistics are outside that interval, and p-value is smaller than 0, 05, H0 will be rejected and it will be stated that the variables have a significant influence over the ranking.

- The second hypothesis H0: The possession is not determinant on a team's success
  - Does the possession has a relevant influence on a team's performance improvement taking into account the ranking as a measure of success?

H0:  $\beta_3=0, p>0, 05$  Possession does not affect ranking

H1:  $\beta_3\neq0, p<0, 05$  Possession affects final position in the ranking

Significance level: 0, 05; Confidence Level: 95 %

Confidence Interval for the H0:  $(-1, 96 < t \text{ or } F < 1, 96)$

In the case that observed F or t-statistics falls into the confidence interval, and p-value is bigger than 0, 05; the H<sub>0</sub> will be accepted and then stated that the possession doesn't have a significant influence over the ranking.

- The third hypothesis: A third hypothesis H<sub>0</sub>: received goals are more determinant over performance than the scored goals
  - Is it true that the received goals have a stronger influence than scored goals over a team's success given the analyzed data?

H<sub>0</sub>:  $\beta_{\text{received}} > \beta_{\text{scored}}$  ; Received Goals have a stronger effect over UCL ranking

H<sub>1</sub>:  $\beta_{\text{received}} < \beta_{\text{scored}}$  ; Scored Goals have a stronger effect over UCL ranking

Significance level: 0, 05; Confidence Level: 95 %

Confidence Interval for the H<sub>0</sub>: (-1, 96 < t or F < 1, 96)

In this case, the ANOVA and the Regression Model coefficients will be contrasted. If p-value of the received goals is smaller than the p-value of the scored goals, in that case the H<sub>0</sub> will be accepted and it will be stated that the received goals have a stronger effect over the ranking than the scored goals. In the opposite case where the p-value of the scored goals is smaller than the p-value of the received goals; the H<sub>0</sub> will be rejected and it will be infer that the scored goals are more significant with respect to a team's success (the final position in the ranking).

## 6. RESULTS

### The descriptive analysis

As it was said before this analysis will permit to have a general picture of every variable's evolution during every phase of the competition in the five years observed period. Next there's a table showing all the data and some graphs with the most relevant variables. This section is being extended in the APPENDIX I (after References Section) that count with more graphs, figures and explanations especially of the regression analysis.

**Table 3 : Descriptive Analysis of UCL using average parameters observed on every phase of the competition**

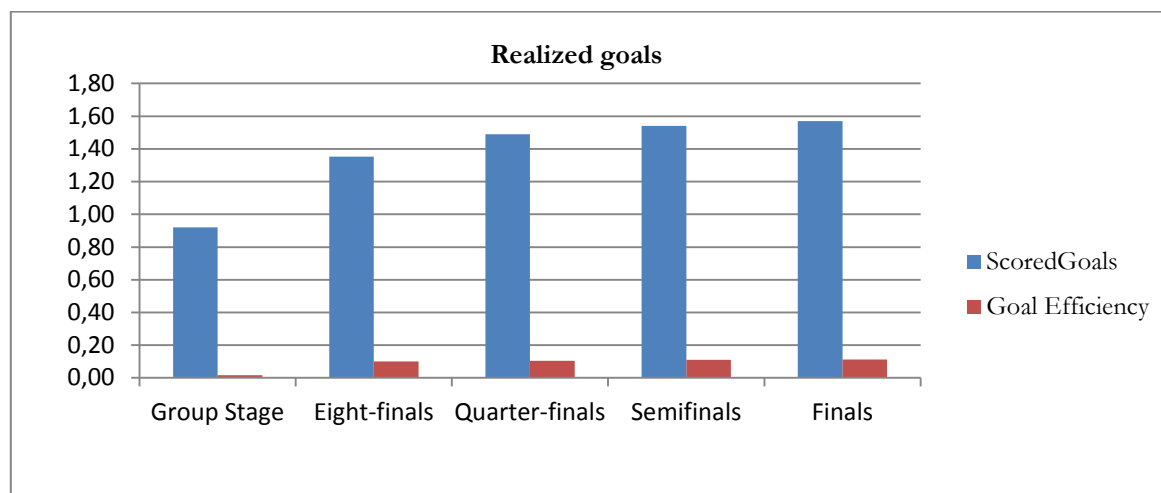
Variable	Group Stage	Eight-finals	Quarter-finals	Semifinals	Finals	UCL Average
ScoredGoals	0,92	1,35	1,49	1,54	1,57	1,53
ScoredGoalsAway	0,38	0,49	0,61	0,61	0,59	0,60
Goal Efficiency	2%	10%	10%	11%	11%	11%
PossessionUefaWebsite	44%	45%	45%	46%	45%	45%
RealizedShots	12,01	13,88	14,24	13,88	14,04	14,05
RealizedAreaCenters	24,09	24,95	25,75	26,27	25,79	25,94
ReceivedGoals	1,68	1,13	0,99	0,91	0,71	0,87
ReceivedGoalsHome	0,67	0,44	0,36	0,31	0,31	0,32
Defensive Weakness	12%	9%	8%	7%	6%	7%
CommittedFouls	16,40	16,91	16,39	15,73	16,32	16,14
LostBalls	72,86	72,65	73,99	76,06	76,06	75,37
RecoveredBalls	49,32	52,11	53,08	52,88	54,31	53,42
ReceivedShots	14,36	12,62	12,49	12,48	12,39	12,46
AreaReceivedCenters	26,16	23,84	24,37	23,33	23,97	23,89
Variable	Group Stage	Eight-finals	Quarter-finals	Semifinals	Finals	UCL Average
ReceivedShotsOnGoal	5,49	4,43	4,38	4,20	4,05	4,21
GoalkeeperSaves	3,73	3,16	3,31	3,21	3,28	3,27
YellowCards	1,87	1,80	1,82	1,74	1,50	1,69
RedCards	0,11	0,09	0,09	0,08	0,05	0,07

Source: Own research; Optasports Database

In **Table 3** it can be observed the average parameters the different profile teams had in UCL during the five years period: the teams that didn't pass the group stage, teams that achieved just the eight or quarter-finals and so on.

This can be useful to make observable conclusions, such as the average number of scored and received goals per match needed to reach an objective phase. An example of a mixed interpretation: if a team scores an average of more than 1, 54 goals/match and receives less than 0, 9 goals/match it has high probabilities to reach the UCL final. Of course these conclusions are not exact, but they can give an idea of a team's needs to reach success and moreover, one could get this ideas before analyzing anything: a team that scores more goals than it receives in every match can win everything.

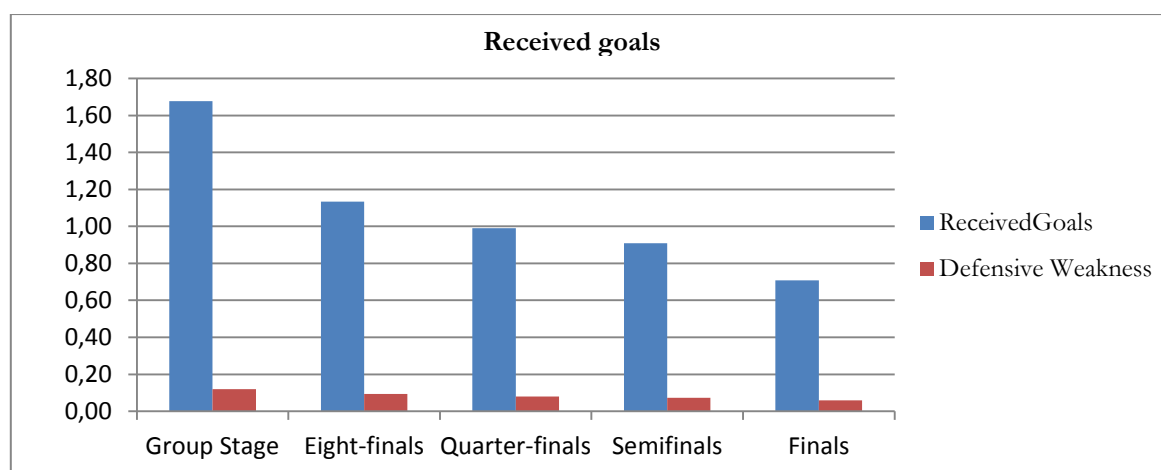
**Graph 1: Evolution of the two most objective attack variables during UCL phases between 2003 – 2008 seasons**



Source: Own research; Optasports Database, Table 3

In **Graph 1** it can be observed an increasing trend on the scored goals; thus it can be said the most successful teams were characterized by having a high goal scoring efficiency. In the other other attack parameters (realized shots, realized centers) the graph follows the same tendency as it can be observed in **Table 3**.

**Graph 2: Evolution of the two most objective defensive variables during UCL phases between 2003-2008 seasons**

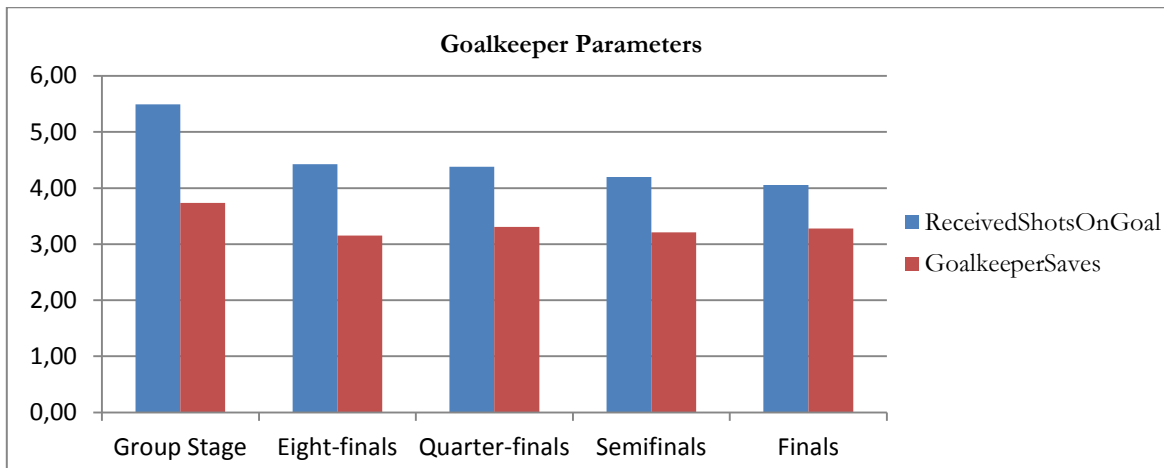


Source: Own research; Optasports Database, Table 3

In this case, the defensive weakness and the received goals are much lower for the teams that reached the advanced stages of the UCL and there's a decreasing trend in the general defensive

parameters over the next tournament phases. This tells us that the teams advancing in the further phase receive less and less goals by phase and have decreasing parameters as well.

**Graph 3: Evolution of the team's goalkeeper parameters during UCL phases between 2003-2008 seasons**



Source: Optasports Database, Own research, Gretl, Table 3

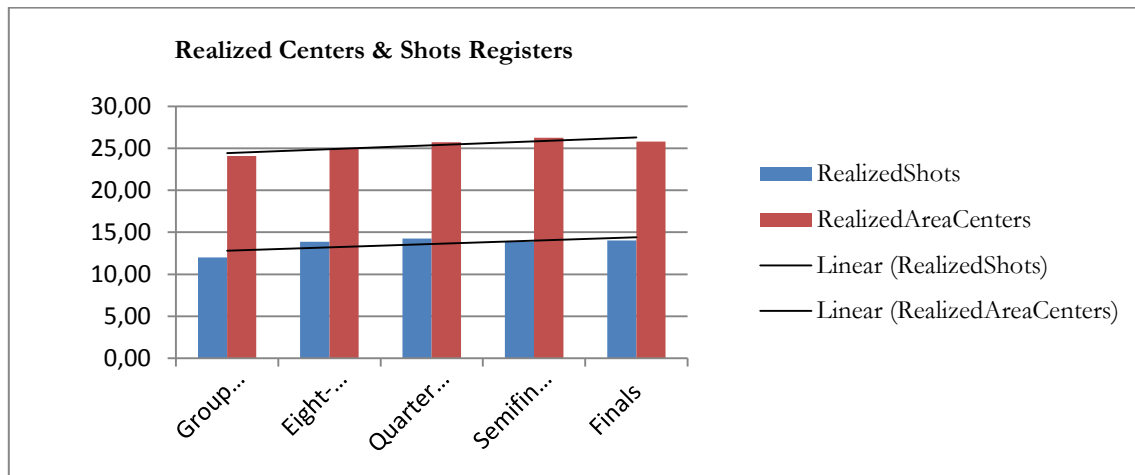
Here it can be observed that the goalkeeper interventions on the teams achieving a further phase were decreasing.

From **Table 3** and **Graphs 1 – 2 - 3**, we can see first that teams that reached the last rounds of the tournament had higher parameters on total scored goals and goal efficiency. There's an increasing trend on those factors starting with the group phase which it may make us think that teams with the highest capacity to score goals made the difference.

Thinking about the defensive parameters it can be observed from the **Graph 2** that the defensive weakness and the received goals values decreased on average as there were observed teams in a more advanced phase. This means that the teams that reached the finals or semifinals, a part of having higher goal efficiency they had moreover a lower defensive weakness which made them being much more efficient on winning and advanced stage by stage in the competition.

The **Graph 3** is also interesting to observe; the goalkeeper parameters over the five tournaments stages use to decrease. This means the role of the goalkeeper in successful teams is less participative than the teams being on a lower phase basically because it receives fewer shots on goal. What about the other technical playing parameters?

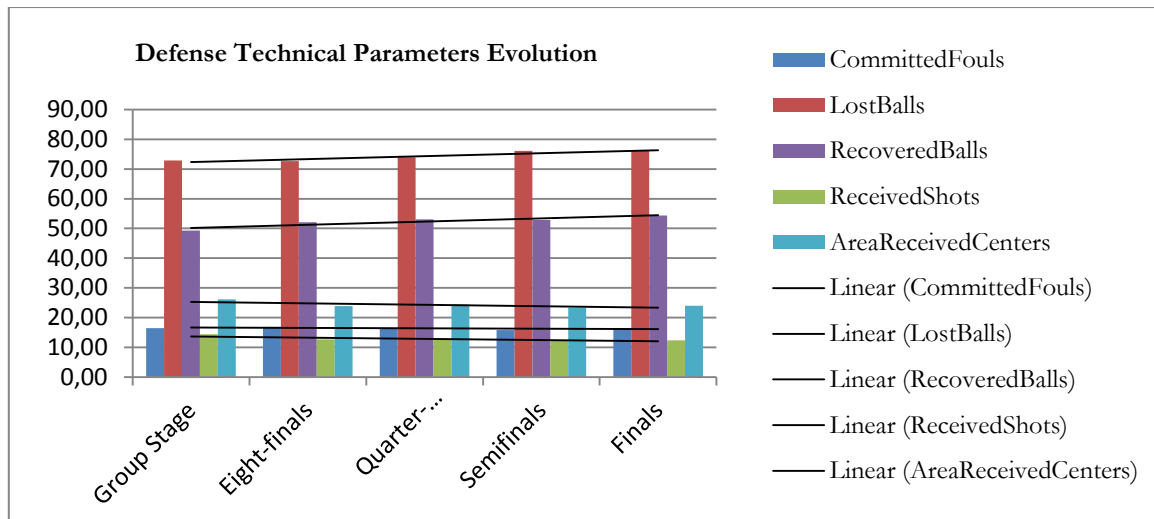
**Graph 4: Attack performance indicators**



Source: Optasports Database, Own research, Gretl, Table 3

In **Graph 5** we can observe an increasing trend on the attack technical parameters; this means that teams that achieved a further phase in the competition were registering higher values for variables such as realized shots and area centers during the observed period.

**Graph 5: Defense performance indicators**



Source: Optasports Database, Own research, Gretl, Table 3

In the **Graph 5** we observe a slow decreasing trend over factors as area received centers, committed fouls and received shots. This is not very clear though and that's why it is needed a statistical analysis to see which variables are actually having a stronger impact on ranking as just by observing this graph it can't be clear. The variables lost balls and recovered balls have an

increasing trend, the teams that achieved a final phase were characterized for losing more balls per match but they were also much efficient when recovering them.

So from the descriptive analysis we got the importance of attack efficiency, the low defensive weakness and the goalkeeper decreasing importance on the successful teams as the competition advances stage by stage. Moreover it was seen that the graphs do not give us a clear estimation of every performance indicator influence over the ranking. That's why more analyses are needed to see which are the most determinant ones.

### The Correlation Analysis

**Table 4: Correlation Matrix, Ranking vs. Independent Variables**

Variable	Correlation Coefficient	Relationship Intensity
Received Goals	0,5861	Medium positive relationship
Received Goals at Home	0,4691	Low Correlation
Received Shots	0,3057	Low Correlation
Received Shots on Goal	0,3057	Low Correlation
Goalkeeper Saves	0,2606	Low Correlation
Received Area Centers	0,1776	Very Low Correlation, there is no evidence of relationship
Red Cards	0,1478	Very Low Correlation, there is no evidence of relationship
Yellow Cards	0,1093	Very Low Correlation, there is no evidence of relationship
Committed Fouls	0,0412	Very Low Correlation, there is no evidence of relationship
Possession Uefa Webpage	-0,0282	Very Low Correlation, there is 0 evidence of relationship
Lost Balls	-0,1055	Very Low Correlation, there is no evidence of relationship
Realized Area Centers	-0,1738	Very Low Correlation, there is no evidence of relationship
Scored Goals Away	-0,2805	Low Correlation
Recovered Balls	-0,3082	Low Correlation
Realized Shots	-0,3333	Low Correlation
Scored Goals	-0,4727	Low Correlation

Source: Optasports Database, Own research, Gretl.

**Table 4** is showing the results for this analysis. We can observe the correlations do not overpass an  $r=0,58$  coefficient, which tells us that variables are not highly correlated to the ranking. A very interesting idea we can get from this analysis is that the possession **R**



correlation coefficient ( $r=0,028$ ) is the lowest one among the other variables together. This is surprising as there are many authors like Lago & Dellal (2010) that affirmed the possessions influence over a team's performance; and here there's a first indication that maybe the possession doesn't have any influence over the ranking.

Later on there will be given more arguments to see if this idea holds or not. Another variable with a low correlation to ranking is the committed fouls ( $r= 0, 041$ ), which can be understandable and interpreted as all teams commit a pretty close number of fouls during the competition. So the outstanding idea from this analysis is that possession does not affect the ranking at all.

**The ANOVA Analysis**  
**Table 5 : Results from the ANOVA analysis**

ANOVA Analysis	P-value	F-statistic
ReceivedGoals	1,28E-08	4,03
ScoredGoals	1,12E-05	2,94
ReceivedGoalsHome	0,0003	2,39
ReceivedShotsOnGoal	0,0011	2,21
ScoredGoalsAway	0,018	1,72
RealizedShots	0,06	1,48
GoalkeeperSaves	0,43	1,02
YellowCards	0,48	0,99
ReceivedShots	0,5	0,97
RedCards	0,56	0,93
CommittedFouls	0,73	0,81
RecoveredBalls	0,73	0,82
AreaReceivedCenters	0,73	0,82
RealizedAreaCenters	0,95	0,59
LostBalls	0,95	0,59
PossessionUefaWebsite	1	0,17

Source: Optasports Database, Own research, Gretl.

The ANOVA parameters were used just to make a comparison to the regression results. There were clear differences, and this analysis found as relevant factors just the goal registers and the realized shots; the other parameters seemed not to have any influence over the ranking. That's why, the Regression and hypothesis testing analysis were preferred to the ANOVA analysis, because it gave us a more reasonable response to the initial hypothesis we had.

From **Table 5** there can be extracted some consistent conclusions that are similar to the previous ones there were just found out in the descriptive and correlation analysis.

The possession parameter has the highest p-value, so in all cases we would accept the null hypothesis; this analysis gives one more argument to state that possession does not affect a team's success in UCL, at least applied to the present case (2003 – 2008 seasons).

At this stage, the most significant variables would be the goals related one plus one technical parameter: received goals ( $p < 0,01$ ); scored goals ( $p < 0,01$ ); received goals at home ( $p < 0,01$ ); received shots on goal ( $p < 0,01$ ) and scored goals away ( $p < 0,02$ ).

From this ANOVA analysis we can infer now that the defensive goal parameters are quite important, having lowest p-values than the attack goal ones. In the third hypothesis, we could infer that the received goals are more important and accept the null hypothesis according to ANOVA results.

Considering the first and the main hypothesis, according to the ANOVA criteria, just one of the parameters would influence a team's success: received shots on goal. The other relevant variables would be the goal ones. That's why, in order to get a more complete idea about the determinant indicators it is needed to do some more analyses

### **The Regression Analysis**

- a. Looking at the **Figure 1 in the Appendix I** we get an adjusted R-square of 0,55 which means that the 12 selected variables can explain at 50% the variations on the ranking so the selected variables can explain the ranking evolution in the 50 % of the cases. That's why, it can be said that there are many other variables influencing the ranking that are not taken into account in this study. But, what about the actual selected performance indicators were selected? Which are their coefficients and which are the ones having the lowest p-values? In the next table there will be shown all the relevant data.

**Table 6: Regression Results from Figure 1 (Appendix I)**

Variables	Betas	T-statistic	P-value	Figure
Received Shots on Goal	6,790	8,548	0,000000	Figure 1
Goalkeeper Saves	-4,522	-4,014	0,000095	Figure 1
Yellow Cards	3,477	3,590	0,000400	Figure 1
Realized Shots	-0,767	-2,926	0,004	Figure 1
Realized Centers	0,370	2,534	0,012	Figure 1
Received Shots	-0,837	-2,463	0,015	Figure 1
Received Centers	0,363	2,305	0,023	Figure 1
Recovered Balls	-0,404	-2,211	0,029	Figure 1
Lost Balls	0,173	1,033	0,303	Figure 1
% of Ball Possession	6,655	0,978	0,330	Figure 1
Red Cards	2,769	0,606	0,546	Figure 1
Committed Fouls	0,119	0,503	0,616	Figure 1

Source: Optasports Database, Own research, Gretl.

From **Table 6**, looking strictly at the p-values we observe the lowest ones and they correspond to 8 technical performance indicators: received shots on goal ( $p < 0,01$ ); goalkeeper saves ( $p < 0,01$ ); yellow cards ( $p < 0,01$ ); realized shots ( $p < 0,01$ ); realized centers ( $p < 0,015$ ); received shots ( $p < 0,02$ ); received centers ( $p < 0,023$ ) and recovered balls ( $p < 0,029$ ).

All these performance indicators should be especially taken into account when training and trying to improve the performance according to this 2003-2008 analysis. Once again, the possession with a 0,3 p-value; is not a determinant variable and it doesn't have a significant influence on a team's performance improvement; as well as the lost balls, red cards and committed fouls indicators.

#### The betas interpretation

In this case, as the ranking goes from 1<sup>st</sup> to 32<sup>nd</sup> position, being the 1<sup>st</sup> position the best one to be it will be observed that the defensive coefficients are positive; being negatives the attack ones. What does that mean? It means that if a parameter has a negative beta coefficient, it helps the team to lower its actual position in the ranking. For example if a team is on the 5<sup>th</sup> position,

and it has a scored goals beta of - 4, it means that in the case the team scores 4 goals it can go one position lower in the final ranking, moving from the 5<sup>th</sup> position to the 4<sup>th</sup> one.

In our case, taking the received shots on goal  $\beta_9=6,79$  as an example. This means that if a team receives more than 6,79 shots on goal per match, it is going to move to a lower position. Obviously there are many other factors influencing a team's success, but given the model it was built on this study, that would be a typical Beta estimator interpretation.

**b.** After performing the established regression analysis with the model from **Table 2** (the one with just 12 game related performance indicators), there was performed a new regression analysis that took into account the goal variables this time. The main reason was to compare the p-values of the regression with the ones obtained in the ANOVA analysis in order to see if they lead us to the same conclusion.

**Table 7: Regression results including goal variables in the first regression model**

Variables	Betas	T-statistic	P-value
Scored Goals	-7	-5	1,06E-06
Received Goals Home	8	4	0,0002
Scored Goals Away	6	2	0,0405
Received Goals	0,737	0,5139	0,608

Source: Optasports Database, Own research, Gretl.

From this table it can be inferred that the scored goals variables have a stronger influence over the ranking as their p-values are lower to the received goals variables. This finding is totally opposed to the ANOVA analysis result, which stated that the defensive indicators were more important. Well, at least, there can be observed the same variables taking into account different analyzing criteria's and the readers can reflect about this idea.

So up to this point, it's a bit unclear if the scored or the received variables are more determinant over a team's success (3<sup>rd</sup> hypothesis).

### The Confidence Intervals Analysis

**a.** The first hypothesis: The 12 performance indicators do not affect the ranking.

“Was it true that the 12 selected variables affect the ranking position in UCL?”

H<sub>0</sub>: Predictor variables do not influence the final position in the ranking

$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12} = 0$

Significance level: 0,05; Confidence Level: 95 %

Confidence Interval for the  $H_0$ :  $(-1,96 < t \text{ or } F < 1,96)$ . Normal Distribution – 2 tails.

Given the fact that there are 8 variables having a lower p-value than the significance level ( $p < 0,05$ ); which moreover have a T-statistic falling out of the confidence interval; the  $H_0$  is rejected at a 5% significance level, and it is concluded that those eight variable have a significant influence over the final position in the UCL ranking.

For example, the goalkeeper saves have a p-value of 0,0000095 causing the rejection of  $H_0$  in the 95% of the cases; its T-statistic=-4,014 falls clearly out of the  $(-1,96; 1,96)$  confidence interval.

The eight influent performance indicators are the received shots on goal, the goalkeeper saves, the yellow cards, the realized shots, the realized centers, the received shots, the received centers and the recovered balls.

This interpretation is based on the Regression performed using the Gretl software with a heteroskedasticity correction, for the 12 performance indicators explained on **Table 6** and **Figure 1** on the Appendix I.

**b.** The second hypothesis  $H_0$ : The possession is not determinant on a team's success.

“Did the possession had a relevant influence on a team's performance improvement taking into account the ranking as a measure of success?”

$H_0: \beta_3=0, p>0,05$  Possession does not affect ranking

Significance level: 0,05; Confidence Level: 95 %.

Confidence Interval for the  $H_0$ :  $(-1,96 < t \text{ or } F < 1,96)$

As it can be observed in the **Table 6** results, the possession has a p-value of 0,3 and its T-statistics is 0,97. As the t-statistic falls into the confidence interval, and it's p-value is bigger than the 0,05 significance level, The  $H_0$  is accepted and it can be said that the possession

didn't have any influence over a team's success in UCL (2003-2008). Using the ANOVA p-value=1 we would sum up one more argument to accept H0.

This result corroborates what (Collet, 2013) affirmed before, that the possession influence in UCL is way lower than in domestic competitions.

c. The third hypothesis H0: received goals are more determinant over performance than the scored goals.

“Was it true that the received goals have a stronger influence than scored goals over a team's success given the analyzed data?”

H0:  $\beta_{\text{received}} > \beta_{\text{scored}}$  ; Received Goals have a stronger effect over UCL ranking

Significance level: 0,05; Confidence Level: 95 %

Confidence Interval for the H0: (-1,96 < t or F < 1,96)

To get any idea for this hypothesis one should check the **Table 5** and **Table 6** which are based on the Regression and ANOVA analysis respectively. **Table 7** has a big difference regarding the goals variables; from the Regression performed analysis with 16 variables (12 technical parameters plus 4 goal variables) it is inferred that the received goals have no influence over the ranking as its p-value is equal to 0,6. In order to make a choice there was calculated an average p-value of all 4 goal variables taking into account both ANOVA and the Regression results. Check **Table 8** on next page.

The third hypothesis (Received Goals have a stronger impact on ranking than scored goals) would be accepted just in the case we consider the scored goals away and the received goals at home. It is true that received goals at home have a stronger influence over a team's performance as its p-value is smaller than the p-value of the variable scored goals away (0,0002 < 0,0405). But if we speak in general, taking the variables received and scored goals in total terms (home and away) it is clear that the scored goals have a stronger influence as they have a smaller average p-value as it can be observed in the next table.

**Table 8: Average p-value of the 4 goal variables**

	Average P-value
Scored Goals	1,4628E-02
Received Goals	0,152150003

Source: Optasports Database, Own research, Gretl.

The scored goals average p-value took into account the total scored goals and the scored goals away average p-value in the Regression and ANOVA results (Table6 and Table5). The received goals average p-value was measure taking into account the received goals and received goals at home p-value results from the same tables. After observing that the average p-value of scored goals is lower than the received goals in average; we can state that the scored goals usually have a stronger impact on a team`s success in UCL (2003-2008). That`s why the H0 is rejected with a 5% risk level.

**Important Hint:** A part of all graphs and tables seen in this results section, there can be consulted some additional ones **Appendix I** section (after the bibliography); these graphs are especially scatters that are helpful to see the linear influence of all variables over the ranking and subsequently a team`s success. There are other graphs and figures as well showing the performed tests to build a reliable regression model, so it is an interesting section to review.

## 7. CONCLUSIONS

The results from the four analyses led us to identify the most determinant performance indicators in a team`s success applied to the 2003-2008 UCL Optasports database. The descriptive analysis (**Table 3**) showed that teams which achieved a further phase in UCL were especially efficient in goal scoring and had high registers for the attack parameters such as realized shots and realized area centers. The possession was an exception and it was not a differentiator factor among the successful teams. We observed that even that teams reached finals had a lower possession average than the ones that reached semifinals (45 % < 46 %) which is an additional argument to think this is not a very determinant factor for winning the competition. In any case it is clear there are teams that still choose to focus on this factor and reached success lately, but they are a rare example not able to be explained by statistics: Fc Barcelona.

The scored goals registers showed a very superior average for the finalists; they had a 70 % more goal efficiency than the teams that couldn't pass the group stage. This can be seen as a fundamental factor when trying to reach performance then, it is the one that originates the highest difference among teams that reached a different phase. The received goals are determinant as well; the average received goals for the finalists are 60 % less than the average received goals for the teams that reached just the group stage. This is the next most determinant factor if we would take into account just the goal variables; and among both of them we see that the scored goals are more decisive (+70 % scoring efficiency > - 60 % defensive weakness).

In the other game related parameters, the effect is a bit more unclear and that's why it was needed a statistical analysis to see which are the most determinant ones because we couldn't get any conclusion just by observing the data; the differences among the parameters were too little to state anything. That way, it was observed that the attack parameters such as realized shots and area centers had increasing values for every phase ; in the opposite side the defensive ones had decreasing values for every phase. The interpretation is that the teams that achieved a final phase had normally higher values on realized shots and realized area centers, and at the same time they had lower values for received shots, received area centers and goalkeeper saves. They were basically being more efficient when attacking and also when defending.

There's an interesting variable to observe, the recovered balls. This is the only defensive variable having an opposite effect compared with the others; it increases for the most successful teams instead of decreasing as the other defense variables; it has a negative relationship to the ranking. If we think about its logic, it is normal to have that effect over the ranking: a team that recovers more balls can start attacking and generate new goal occasions, then the ranking moves down as a team's position improves (**Figure 3**). Observing the goalkeeper parameters, it can be said that the finalists, being the most successful teams in the competition, had much better results over this parameters as well: they had a 30 % less received goals on average than the group stage teams, and a 12 % less goalkeeper interventions. That's why, their defensive efficiency it was clearly important as well as the offensive one; they could reach finals by having less received shots on goal and registered lower average goalkeeper interventions.



The correlation analysis helped us mainly to see that the possession had one of the lowest correlation coefficient to a team's success; an idea we observed before in the description analysis as well. This was one more argument to state that the possession doesn't have a clear effect on status when we speak about all the participating teams in the UCL.

Moving to the regression results, there were found eight technical variables that could affect a team's success. Coaches, anyone interested or competing in UCL should pay attention to these indicators: realized shots( $p < 0,01$ ); realized centers( $p < 0,015$ ); recovered balls( $p < 0,01$ ); received shots( $p < 0,02$ ); received centers( $p < 0,03$ ); received shots on goal( $p < 0,01$ ); goalkeeper saves( $p < 0,01$ ) and yellow cards( $p < 0,01$ ). These parameters are important because they have the lowest p-values, meaning they can make a difference when competing in UCL. We knew that goal efficiency was going to be decisive, but if the aim of this study was to identify the technical performance indicators related just to the way of playing, those are the playing factors that most affect a team's performance. The fact of having the lowest p-values made us to reject the initial hypothesis that "the selected parameters didn't affect a team's success" and that's why they are the most decisive ones.

It was previously seen that there were some contrary findings over the possession. From the description, correlation and regression analysis, this study stated that the possession doesn't have a relevant influence in football performance at least in UCL as its p-value was around 0,3 being one of the highest ones. In the correlation analysis the correlation coefficient was close to zero; In **Figure 3** in the Appendix we observe the fitted graph between possession and UCL ranking and there's no relationship at all representing a quite disperse data. Those arguments are enough to consider that the possession doesn't actually have a clear influence over a team's success and other factors should be prioritized.

Looking at the ANOVA goal variables analysis (**Table 8**), it can be stated, that a team that focuses to improve the goal efficiency and the scoring goal rates it will surely be able to reach more success than a defensive focused team. The reason to state that is that the scoring goals have a lower p-value and that would mean a more decisive impact over a team's success. In any case, the defensive strength is important as well and it is the second most decisive factor.

If there would be made a top of the most important overall variables, it would be like this:

1. The goal efficiency : scored goals
2. The defensive strength : received goals
3. Game Performance Indicators : attack and defensive ones (the 10 included in the Regression Model)
4. Referee Parameters: yellow and red cards.

The reasons of this top are based on the p-values of every variable with respect to a team's success observed in the previous regression, ANOVA and the descriptive analysis. This way, anyone interested in football analysis can make a priority list with variables that have to be taken into account when trying to improve a team's performance. It can be observed that the analysis of the performance indicators done in this study would be the third most important after observing the goal efficiency and the defense strength of a team. Making a general reflection, one can understand why football teams spend more for acquiring attack/scorer players as they seem to be the most important contributors to a team's success in the case the defense and the other positions are at the same level. Thinking about the goalkeeper role on a team's success, it can be said it is a differentiator factor, but counting with a good goalkeeper has a lower influence than having a high scoring player.

Moreover, the possession debate in UCL is clear, this factor does not have a clear influence on a team's performance and there were many proofs during this study that state that; in that case it can be inferred that the counter attack style, or at least the ball recovery indicator is more important or representative and it has a higher impact on a team's performance. These findings about possession coincide with the ones in the Collet's (2013) study.

Making a reference to Lago-Peñas et al. (2011) study, this paper coincides on the shots on goal significance as being one of the important distinctive variables to reach success. Even so, the same findings do not coincide about the possession importance as it was many times commented before: the results of this short study show clear evidences that possession is not a representative and significant factor to reach success in UCL. In my opinion this is the pure statistical interpretation and that doesn't mean that it applies to all cases and circumstances. In recent years we saw excellent teams playing with a possession style such as FC Barcelona or Arsenal and their evolution was opposite: Barcelona won 2 UCL in past 5 years and Arsenal none. But what the findings of this paper say is that after observing the whole group of teams

participating in UCL from 2003 to 2008 there were observed just 8 significant variables which didn't include possession , red cards or committed fouls. Coaches could have a look at these results and make a priorities list to train and try to improve their team's performance by orientating their trainings to the most determinant playing variables. It is known that there could be many other variables that could influence a team's performance in UCL but regarding team playing indicators we saw this study gave some interesting indications about which factors should be taken into account most.

## 8. REFERENCES

- Ashby, K. (2013, 05 28). *UEFA Champions League Final News*. Retrieved 05 17, 2014, from UEFA Champions League Web site:  
<http://www.uefa.com/uefachampionsleague/news/newsid=1957523.html>
- Associated Press. (2011, 07 12). *ESPN Soccer News*. Retrieved 05 17, 2014, from ESPN:  
[http://espn.go.com/sports/soccer/news/\\_/id/6758280/least-1-billion-saw-part-2010-world-cup-final](http://espn.go.com/sports/soccer/news/_/id/6758280/least-1-billion-saw-part-2010-world-cup-final)
- Atkinson, S., Stanley, L., & Tschirhart, J. (1988). Revenue sharing as an incentive in an agency problem: an example from the national football league. *Rand Journal of Economics*, 19, 27 - 43.
- Bate, R. (1988). Football chance: tactics and strategy. In T. Reilly, A. Lees, K. Davids, & W. Murphy, *Science and football* (pp. 293-301). London: E & FN Spon.
- Carling, C., Bloomfield, J., & al., e. (2008). The role of motion analysis in elite soccer: Contemporary performance measurement techniques and work rate data. *Sports Medicine*, 38, 839-862.
- Collet, C. (2013). The Possession game? A comparative analysis of ball retention and team success in European and international football, 2007-2010. *Journal of Sports Sciences*, 31, 123-136.
- Di Salvo, V., Baron, R., Tschan, H., Calderon Montero, F., Bachl, N., & Rigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. *International Journal of Sports Medicine*, 28, 222-227.
- Dobson, S., & Goddard, J. (2011). *The Economics of Football*. Cambridge University Press.
- FIFA. (2014, 05). *About FIFA : Organization - Fifa.com*. Retrieved 05 15, 2014, from FIFA web site:  
<http://www.fifa.com/aboutfifa/organisation/index.html>
- Hughes, M., & Bartlett, R. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences*, 20, 739-754.
- Hughes, M., & Churchill, S. (2005). Attacking profiles of successful and unsuccessful team in Copa America 2001. In T. Reilly, J. Cabri, & D. Araújo, *Science and Football*, 5 (pp. 219-224). London and New York: Routledge.
- Kuypers, T. (1995). The beautiful game? An econometric study of why people watch English football. *University College London Discussion Paper 96/01*.
- Lago-Ballesteros, J., & Lago-Peñas, C. (2010). Performance in team sports: Identifying the keys to success in soccer. *Jour*

*nal of Human Kinetics*, 25, 85-91.

Lago-Peñas, C., Lago-Ballesteros, J., & Rey, E. (2011). Differences in Performance Indicators between winning and losing teams in the UEFA Champions League. *Journal of Human Kinetics*, 27, 135 - 146.

Lago-Peñas, C., Rey, E., Lago-Ballesteros, J., Casais, L., & Dominguez, E. (2009). Analysis of work-rate in soccer according to playing positions. *International Journal of Performance Analysis in Sport*, 9, 218-227.

Newbold, Carlson, & Thorne. (2010). *Statistics for Business and Economics*, 7th Global Edition. Pearson.

Palacios-Huerta, I. (2004). Structural changes during a century of the world's most popular sport. *Statistical Methods and Applications*, 13, 241-258.

Peel, D., & Thomas, D. (1992). The demand for football: some evidence on outcome uncertainty. *Empirical Economics*, 17, 323 - 331.

Redwood-Brown, A., Bussell, C., & Singh, H. (2012). The impact of different standards of opponents on observed player performance in the English Premier League. *Journal of Human Sport and Exercise*, 7, 341-355.

Rumpf, M. (2014). *Special Topics, Performance Analysis. Football Science*. Retrieved 05 19, 2014, from Football Science: <http://www.footballscience.net/special-topics/performance-analysis/>

Stanhope, J. (2001). An investigation into possession with respect to time in the Soccer World Cup 1994. In M. Hughes, *Notational Analysis of Sport*, 3 (pp. 155-162). Cardiff: UWIC.

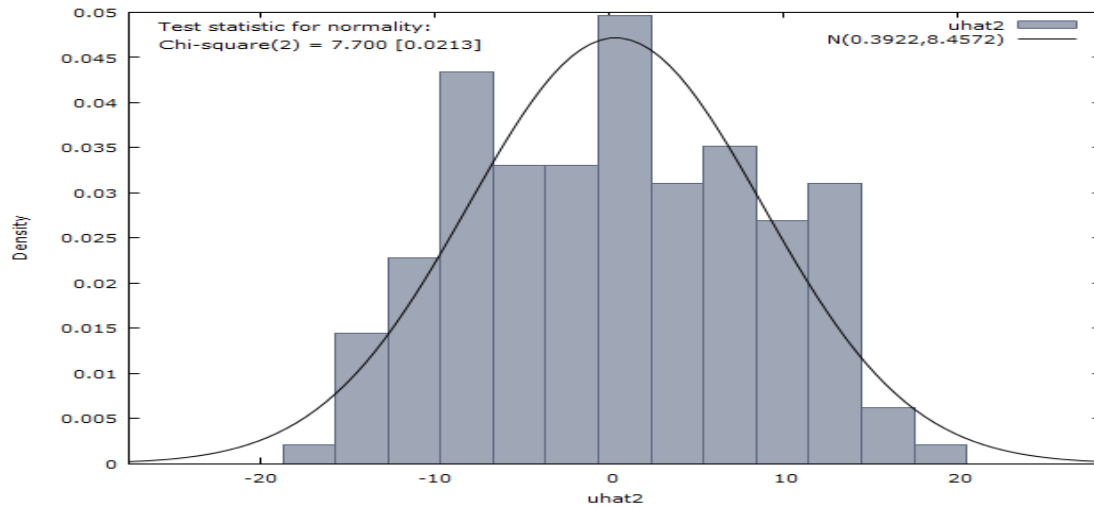
Svensson, M., & Drust, B. (2005). Testing soccer players. *Journal of Sports Sciences*, 23, 601-618.

UEFA. (2014). *UEFA Historical Statistics : 2003/2004 Overall Club Ranking*. Retrieved 05 20, 2014, from UEFA: <http://es.uefa.com/uefachampionsleague/season=2003/statistics/round=1712/clubs/index.html>

## APPENDIX I

### A. Regression Assumptions

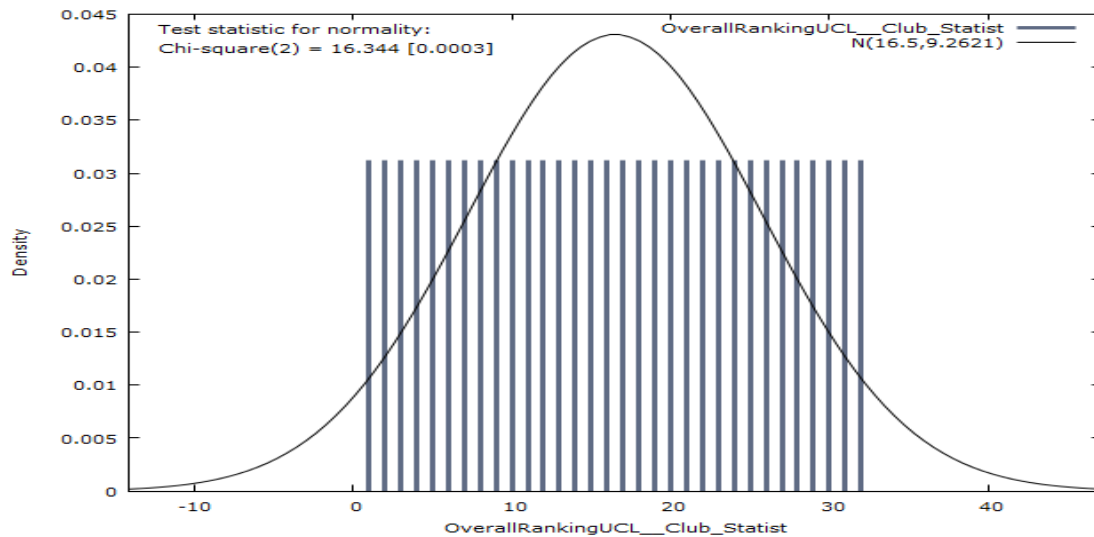
#### Condition 1: Normality Test of Residual of the 12 Variables Regression Model



Source: Own research; Gretl results based on Optasports Database

This is the first proof to see if the regression model works.

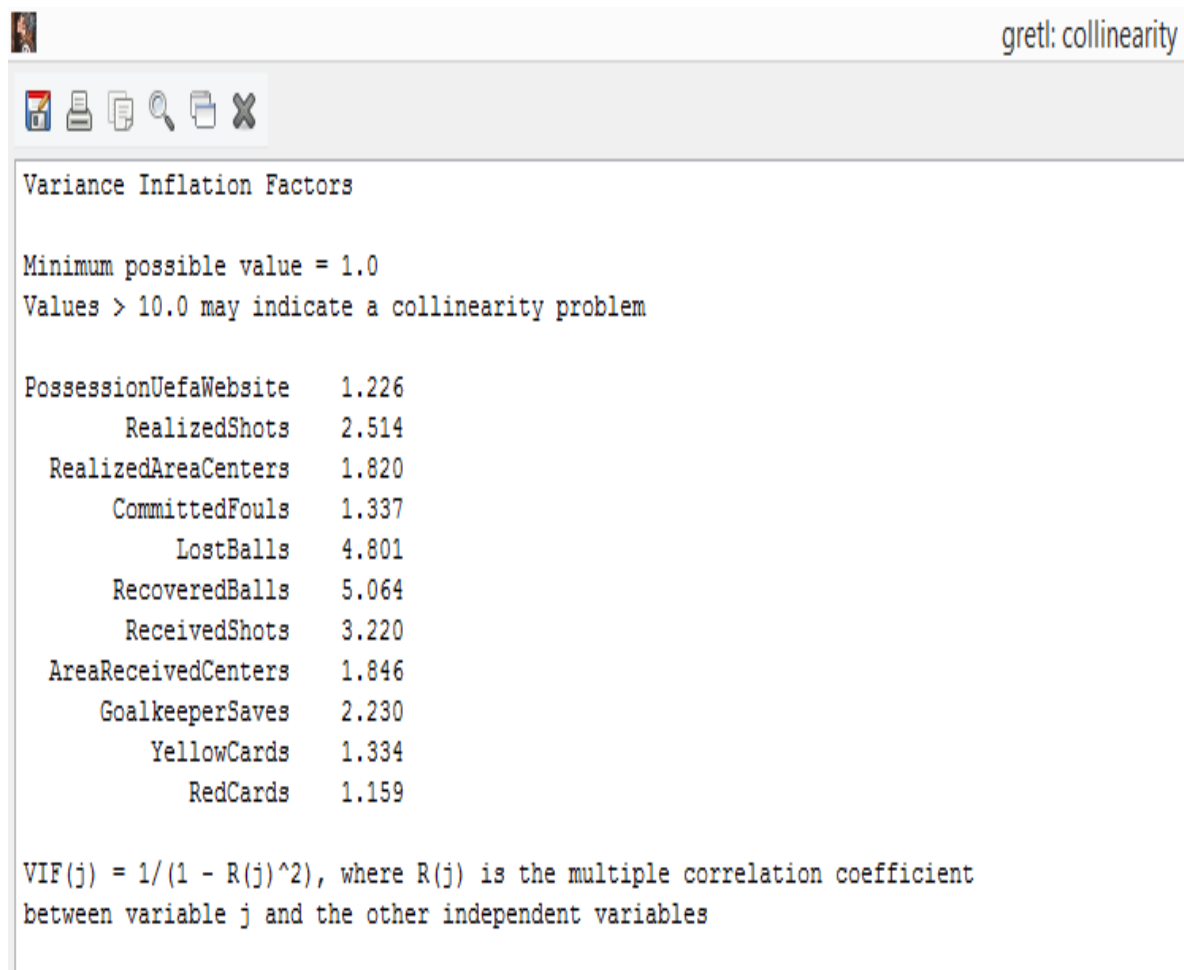
#### Condition 2: Normality Test of the Dependent Variable: UEFA Ranking



Source: Own research; Gretl results based on Optasports Database

A part of this dependent variable normality test there was made a successful normality test for all independent variables as well.

### Condition 3: Linear Independence among Predictor Variables



PossessionUefaWebsite	1.226
RealizedShots	2.514
RealizedAreaCenters	1.820
CommittedFouls	1.337
LostBalls	4.801
RecoveredBalls	5.064
ReceivedShots	3.220
AreaReceivedCenters	1.846
GoalkeeperSaves	2.230
YellowCards	1.334
RedCards	1.159

VIF(j) =  $1/(1 - R(j)^2)$ , where R(j) is the multiple correlation coefficient between variable j and the other independent variables

Source: Own research; Gretl results based on Optasports Database

Condition three is one of the last assumptions that let us confirm the created model was reliable. As there were no values bigger than 10, there were no collinearity problems among the 12 predictor variables.

One of the reasons of excluding the goal and the number of played games from the regression was the collinearity assumption: these variables had a strong correlation with the other independent variables. That's why, in order to obtain the p-value of goal variables there was performed an ANOVA analysis in order not to affect the regression coefficients and the conclusions made out of it.

Figure 1: Regression Results, Regression Model based on Table 2

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

Model 1: Heteroskedasticity-corrected, using observations 1-160  
Dependent variable: OverallRankingUCL\_\_Club\_Statist

	coefficient	std. error	t-ratio	p-value	
-----	-----	-----	-----	-----	
const	-2.21022	8.95986	-0.2467	0.8055	
PossessionUefaWe~	6.65502	6.80796	0.9775	0.3299	
TotalRealizedSho~	-0.767407	0.262276	-2.926	0.0040	***
RealizedAreaCent~	0.370348	0.146132	2.534	0.0123	**
CommittedFouls	0.118970	0.236583	0.5029	0.6158	
LostBalls	0.172976	0.167385	1.033	0.3031	
RecoveredBalls	-0.403621	0.182536	-2.211	0.0286	**
ReceivedShots	-0.836789	0.339749	-2.463	0.0149	**
AreaReceivedCent~	0.362802	0.157414	2.305	0.0226	**
ReceivedShotsOnG~	6.78989	0.794336	8.548	1.48e-014	***
GoalkeeperSaves	-4.52216	1.12654	-4.014	9.47e-05	***
YellowCards	3.47689	0.968400	3.590	0.0004	***
RedCards	2.76916	4.57195	0.6057	0.5457	

Statistics based on the weighted data:

Sum squared resid	449.3560	S.E. of regression	1.748383
R-squared	0.589317	Adjusted R-squared	0.555792
F(12, 147)	17.57834	P-value(F)	6.37e-23
Log-likelihood	-309.6415	Akaike criterion	645.2830
Schwarz criterion	685.2603	Hannan-Quinn	661.5164

Statistics based on the original data:

Mean dependent var	16.50000	S.D. dependent var	9.262082
Sum squared resid	8142.213	S.E. of regression	7.442393

Excluding the constant, p-value was highest for variable 10 (CommittedFouls)

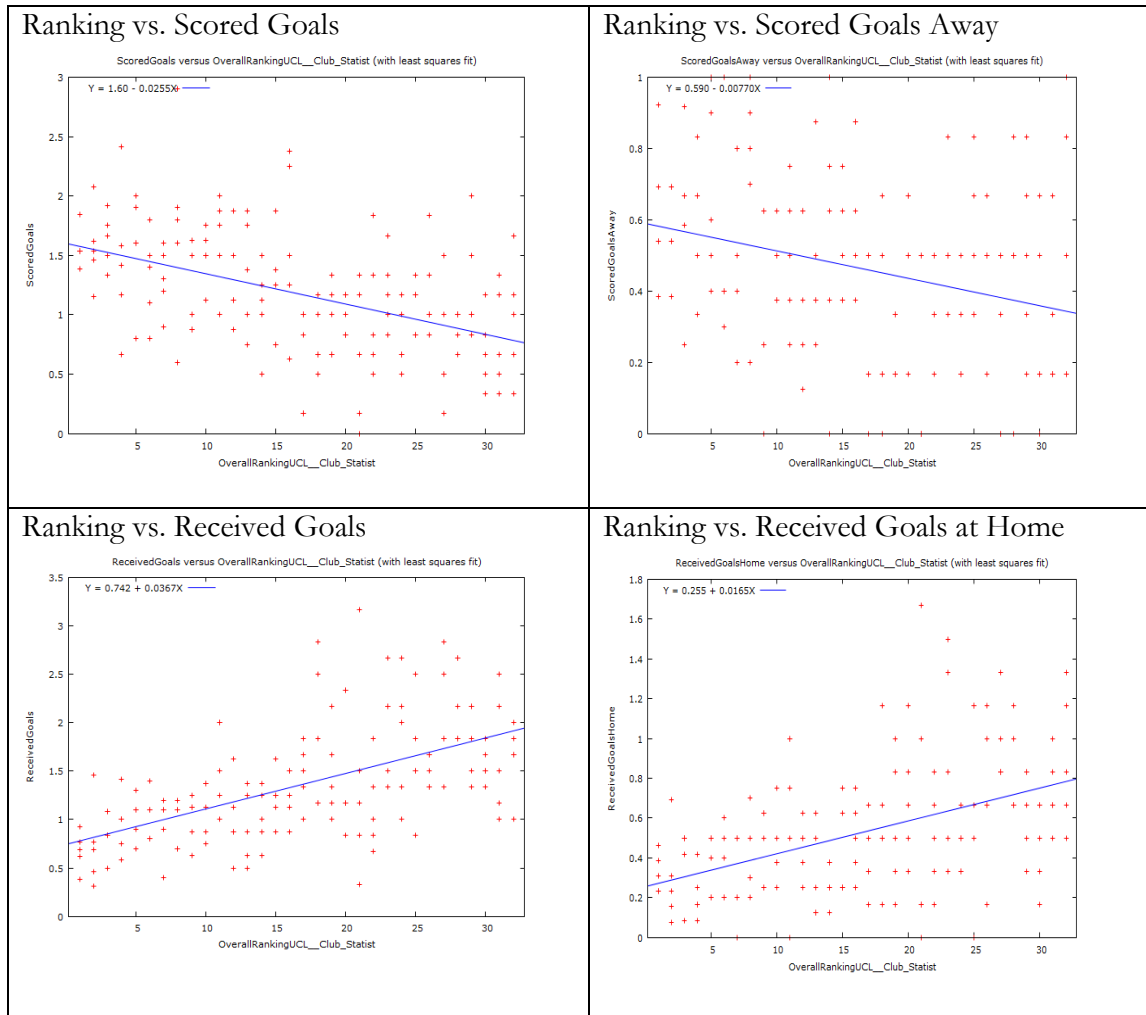
Source: Optasports Database, Own research, Gretl.

We observe here a corrected R-square value of 0, 55, which states that the build regression model could explain just 55 % of the ranking variability. There was build an additional regression for the same 12 variables, but including the goal and number of played games. The new model results improved up to an R-square of 0, 87; which would mean a more complete model than the one we had used in this study. Nevertheless, in that case, the conclusions that could be extracted were irrelevant: the most determinant variables were the number of played games and the goals parameters. As that interpretation wouldn't tell us much about which game parameters are the most influential on a team's performance, this study renounced to include goal and number of played games indicators in the regression model for the main objective of the study; they were included just in the ANOVA and the descriptive analysis.

Despite of having the opportunity to increase the R-square by including more variables; this study was focused to analyze just the relevant playing parameters.



**Figure 2: Linear Relationship, Ranking vs. Goal Variables**



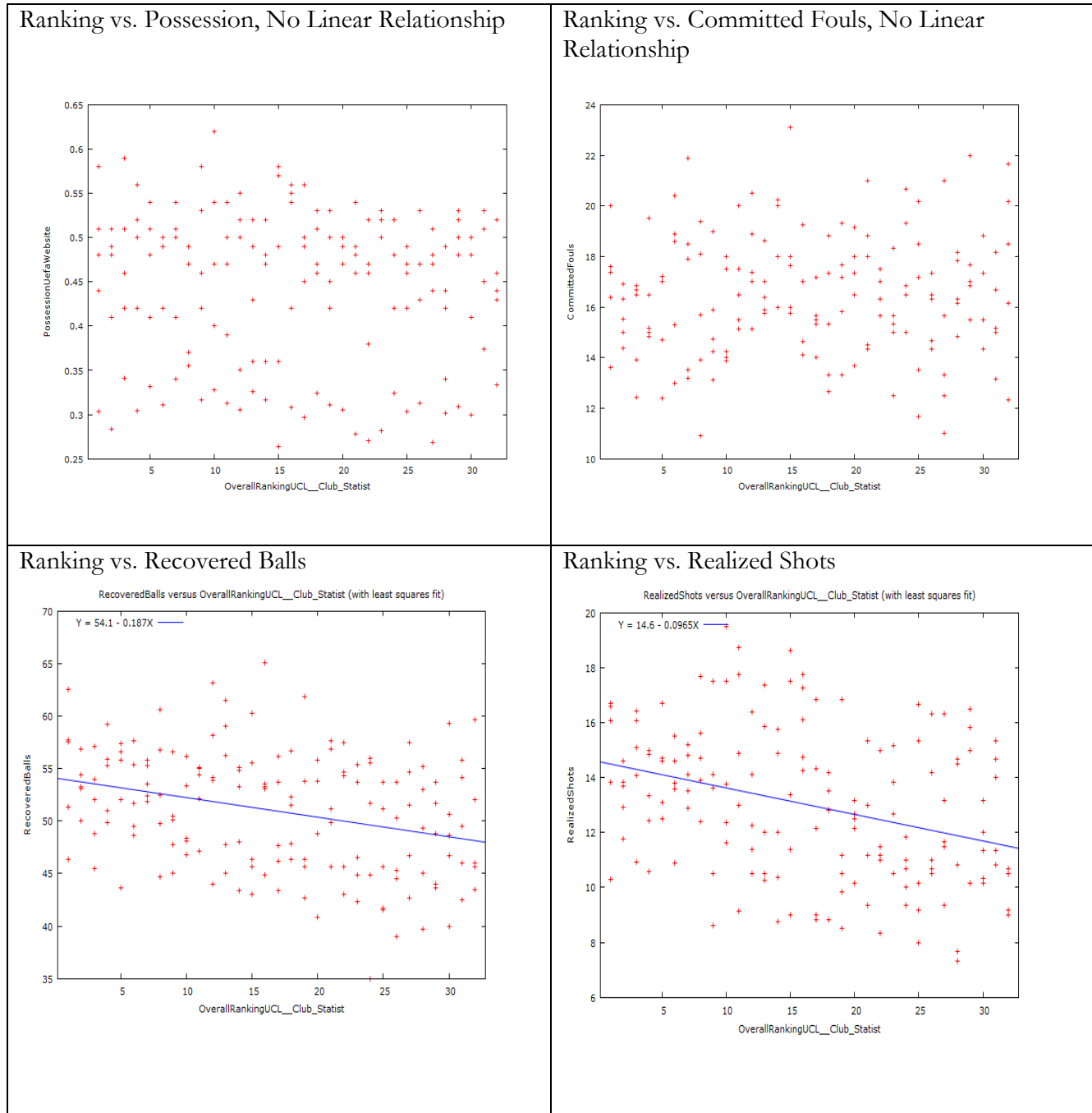
Source: Optasports Database, Own research, Gretl.

The intention of these four scatter graphs was to show their relationship intensity to the ranking by observing the slope of every regression line individually. We observe that the scored and the received goals have a logical relationship to the ranking: if there are more scored goals, a team's position will decrease on a 32-1 scale, which means an improvement on the previous position. On the other side, if it receives more goals, the ranking will move downward and its ranking position will increase on a 1-32 scale. On the horizontal axis there is represented the ranking and on the vertical one there is represented one performance parameter.

What it may be highlighted from **Figure 2** is the fact that the possession has no influence over the ranking as we can observe how dispersed its data is from the linear regression model. We can see there were examples of teams having a low possession and achieved the UCL final or others with a high possession that did not passed the group stage; in the end it seems it's not a relevant factor when competing in UCL. (**Table 3**)

Another interesting fact is that the only defensive parameter affecting the ranking in a positive way is the recovered goals one. When this parameter increases a team's position tends to improve its position in the ranking. (**Figure 3**)

**Figure 3: Linear Relationship, Ranking vs. Non Influential and Influential Parameters difference**



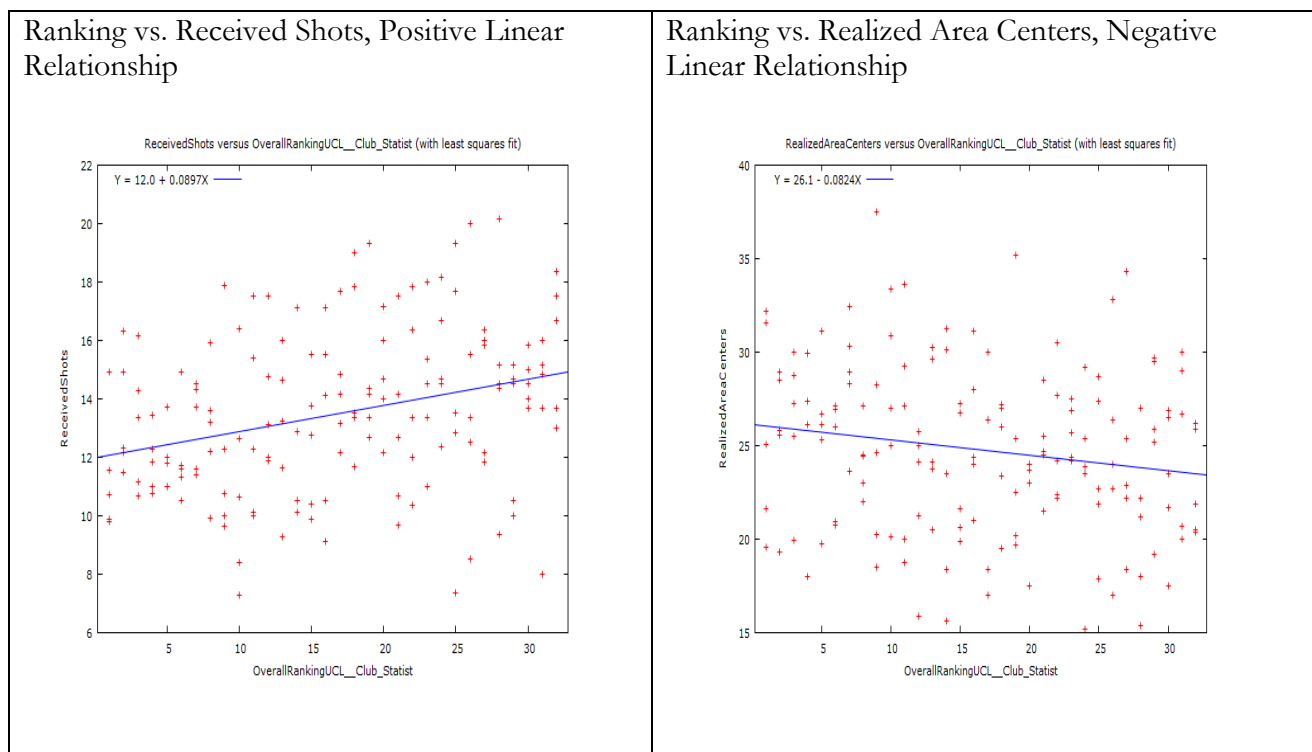
Source: Optasports Database, Own research, Gretl.

On the other hand, the attack factors as realized shots or area centers seem to have a clear “negative” influence over the ranking; when the indicators go up the ranking position moves down which is understood as an improvement in a team’s performance. The defensive factors have a positive linear increasing influence over the ranking as it can be seen in the **Figure 3.**; when the values of these parameters increase, the ranking increases as well meaning that a team position is moving downward from the top: if there are more and more received shots on goal it is normal, lose more matches and move down on the ranking.

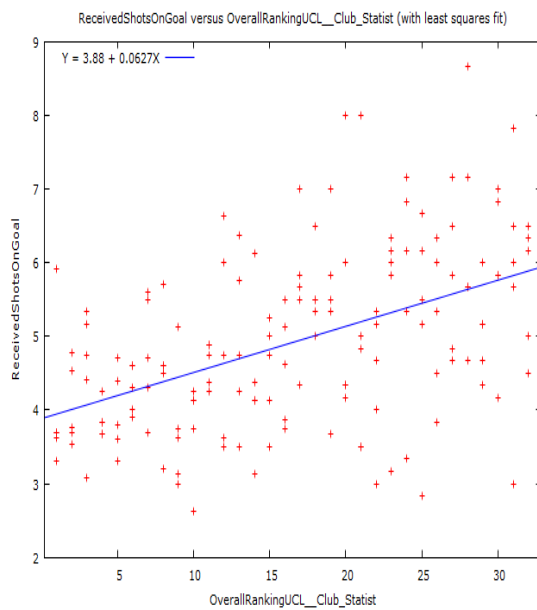
The **Figure 4** show us the difference between an attack and defense variable. The defense parameter (received shots) has a positive relationship to ranking, which is translated in a decline in a team’s position from the 1<sup>st</sup> one to a lower one. The attack parameter (realized area centers) has a negative impact on ranking, when there are more realized area centers a team improves its position in the overall ranking from a 2<sup>nd</sup> position to the 1<sup>st</sup> one for example.

The explanation to this is that the ranking goes from 1 to 32, and an improvement in the ranking is realized when moving from an upper position to a lower one, it is a negative movement (movement from 5<sup>th</sup> to 4<sup>th</sup> position). On the contrary, a decline in the ranking is interpreted on the opposite way, when a team moves from a top position to a lower one (1<sup>st</sup> to 2<sup>nd</sup> position) and that’s why there’s a positive effect on the graphs presented in this section.

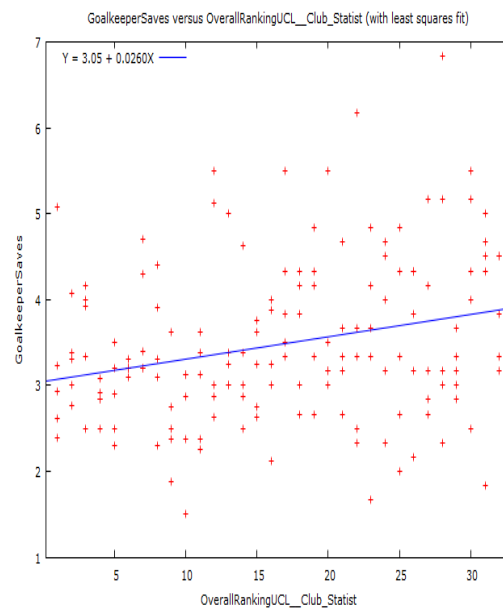
**Figure 4: Linear Relationship Ranking vs. Other Influential Parameters**



### Ranking vs. Received Shots on Goal, Positive Linear Relationship



### Ranking vs. Goalkeeper Saves, Positive Linear Relationship



Source: Optasports Database, Own research, Gretl.

In these last two graphs it can be observed the relationship of the goalkeeper parameters over the ranking. We can say the received shots on goal are more determinant as it has a steeper slope and that is meant to have a stronger impact over a team's success.

## APPENDIX II

Tabla de Competencias Genéricas desarrolladas a la hora de elaborar este trabajo

	Competencia	¿Cómo se adquirió? ¿En qué apartado del trabajo o momento se adquirió?
CB2	Argumentación y resolución de problemas dentro de la rama de estudios	En toda la Realización del trabajo
CB3	Reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética	Creación de un Base de datos , Revisión Literaria, Elección de las variables relevantes para llegar a unas conclusiones
CB4	Transmitir información, ideas, problemas y soluciones tanto a un público especializado como no especializado	Análisis estadístico y Conclusiones
CG01	Capacidad de análisis y síntesis	En el análisis estadístico, los resultados y las conclusiones
CG02	Capacidad de organización y planificación	En todo el proceso de elaboración del TFG
CG04	Comunicación oral y escrita en una lengua extranjera	Inglés
CG06	Habilidad para analizar y buscar información proveniente de fuentes diversas	Revisión Literaria
CG14	Capacidad crítica y autocrítica	Revisión Literaria e interpretación de los resultados
CG15	Compromiso ético en el trabajo	En todo el trabajo , el objetivo de mejorar el conocimiento de otros interesados sobre el mismo tema
CG16	Capacidad en trabajar en entornos de presión	Durante la realización de todo el trabajo
CG17	Capacidad de aprendizaje autónomo	En toda la realización del trabajo trabaje individualmente
CG19	Creatividad	Elección de las Hipótesis

Tabla de Competencias Específicas desarrolladas a la hora de elaborar este trabajo

	Competencia	¿Cómo se adquirió? ¿En qué momento?
CE02	Identificar las fuentes de información relevante y su contenido	Revisión literaria de artículos y estudios previos relacionados con el tema de este trabajo